Skorokhod Embeddings: Non-Centred Target Distributions, Diffusions and Minimality

submitted by

Alexander Matthew Gordon Cox

for the degree of Ph. D.

of the

University of Bath

2004

Signature of Author ....

Alexander Matthew Gordon Cox

Summary

A Skorokhod embedding is a stopping time of a stochastic process such that the stopped process has a given distribution. We consider the problem of finding embeddings when the underlying process is a Brownian motion in one dimension. We are interested in solving the problem for distributions which are not centred. We begin by extending a solution of Perkins from the centred case to any (not necessarily integrable) measure, and demonstrate that this solution maintains a desirable optimality property concerning the distributions of its maximum and minimum.

We then consider the problem of embedding integrable, but not necessarily centred distributions. In the centred case there exists a natural condition on the class of stopping times to determine which stopping times are 'suitable'. In the non-centred case we propose that the class of minimal stopping times is the correct class to consider. We are able to provide simple necessary and sufficient conditions for a stopping time to be minimal. We also demonstrate that the famous embedding of Azema and Yor can be extended naturally to non-centred target distributions and maintains its optimality properties in the class of minimal embeddings.

Finally we consider the case where the Brownian motion starts in a given distribution, rather than just at a single point. We show that techniques of Chacon and Walsh can be extended to the more general case where the means do not agree. In this new setting we prove new equivalent conditions to minimality. We are able to give simple graphical conditions for a stopping time constructed using the Chacon-Walsh technique to be minimal. Further we show that there is a simple interpretation of several known constructions in this framework — the Azema-Yor and Vallois embeddings.

Acknowledgments

I am of course indebted to a number of people for their support and friendship while writing this thesis:

- David Hobson for all his excellent advice and many helpful suggestions, and for introducing me to such an interesting problem.
- All the staff and postgraduates of the Department of Mathematics, who have been unfailingly helpful and friendly.
- The EPSRC for three years of financial support.
- All the friends I have made in Bath, for the occasional discussions and the odd beer, especially Pascal Vogt, Andy Holt, Steve White, Mark Gittoes, Rob Hardy and John Harris.
- All those who have joined me on the football or cricket pitch, and all those who came to PG-TIPS meetings.
- And of course my parents and family, for their love and support throughout my education.

Contents

List of Figures

Chapter 1

Introduction

At the heart of the modern study of Probability theory lies — almost 100 years after Wiener showed its existence — the Brownian motion. It remains one of those rare mathematical objects which is simple enough to describe in a few lines, and yet on closer inspection reveals myriad surprising properties and appears in many real world applications, ranging from finance to biology and queuing theory. It is this connection with applications and the natural questions that arise from considering processes evolving in time that really make the study of such processes a topic distinct from the measure theory that forms the basis of the subject.

Our interest lies within one particular question related to the study of Brownian motion. The question was one first posed by Skorokhod (1965) and has henceforth been known as *the Skorokhod Embedding problem*:

Key Question. Suppose B is a one-dimensional Brownian motion and μ is a distribution on $\mathbb{R}$. When can we find a stopping time T such that B_T has distribution μ?

1.1 Basic Solutions

It turns out that the problem is easy to solve, and we are able to give two easy solutions to the problem immediately. In this introduction we shall leave the details out, and we refer a curious reader to the Appendix for the calculations.

Example 1.1 (A Quick and Dirty solution)**.** The Skorokhod embedding problem is solved by the stopping time

$$T_Q = \inf\{t \geq 1 : B_t = F^{-1}(\Phi(B_1))\}$$

where F is the cumulative distribution of μ and Φ the cumulative distribution of a $N(0,1)$ random variable. It is easy to check that $F^{-1}(\Phi(B_1))$ has the required distribution.

Example 1.2 (Skorokhod's solution)**.** The second solution we present is the original solution due to Skorokhod (1965). He makes the natural assumption that the target distribution is centred. The embedding is defined to be

$$T_S = \inf\{t \geq 0 : B_t \notin [X, Y]\} \tag{1.1}$$

where X and Y are random variables independent of B with a given joint law $\nu \in \mathcal{B}(\mathbb{R} \times \mathbb{R})$ dependent on μ (see Proposition A.2).

So we have answered our original question and also shown that the solution is not unique. This suggests a new question: in given situations when is a particular embedding 'better', and why have I called the first embedding a 'quick and dirty' solution?

The key observation is the following: when the target distribution is centred with a finite second moment we may calculate $\mathbb{E}(T)$ for both embeddings. When we do this we find that for the first example $\mathbb{E}(T_Q) = \infty$ (unless the target distribution is the $\mathcal{N}(0,1)$ distribution) while the second embedding has $\mathbb{E}(T_S) = \mathbb{E}(B_{T_S}^2) < \infty$. We will see shortly that in some of the key applications of Skorokhod embeddings it will be necessary to keep $\mathbb{E}T$ small. So the next question we might ask is: can we find an embedding of μ for which $\mathbb{E}T < \mathbb{E}(B_T^2)$?

The answer to this question lies in the following Lemma (see also Section A.1.3 for a proof):

Lemma 1.3 (Wald's Lemma)**.** *If T is a stopping time of a Brownian motion $(B_t)_{t\geq 0}$ with $B_0 = 0$ such that $\mathbb{E}T < \infty$ then*

(i) $\mathbb{E}B_T = 0$*;*

(ii) $\mathbb{E}B_T^2 = \mathbb{E}T$*.*

So we can't do any better than $\mathbb{E}(B_T^2)$, we can only do a lot worse! In fact, as a consequence of the lemma we can conclude that $\mathbb{E}T < \infty$ if and only if the process $B_{t\wedge T}$ is a $\mathcal{L}^2$-martingale.

Much of the subsequent work will be primarily concerned with the consequences of relaxing the conditions on μ. The case where μ is centred and in $\mathcal{L}^2$ is closely related to the study of $\mathcal{L}^2$-martingales, and when we drop the assumption that $\mu \in \mathcal{L}^2$ naturally we can no longer find a suitable martingale. However for centred target distributions we can relax the criterion to: $B_{t\wedge T}$ is a UI martingale. We will often call such an embedding a UI embedding. For the embedding to be UI we must therefore have $\mathbb{E}B_T = 0$. This acts in much the same way as the condition $\mathbb{E}T < \infty$ and will be satisfied by T_S but not in general by T_Q. This is a partial improvement, but still leaves us with the following question:

Key Question. Let μ be any probability measure on $\mathbb{R}$. Can we find a general class of embeddings which includes 'nice' stopping times such as Example 1.2 but excludes 'nasty' examples like Example 1.1?

We note that the embedding of Example 1.1 will work for any probability measure μ, and in fact it can be generalised easily (using an independent random variable with distribution μ) to work for any recurrent process on any space. The embedding of Example 1.2 can also be generalised for Brownian motion by allowing X or Y to take the value ∞.

1.2 Further Questions

The main emphasis in this work shall be examining the embedding problem when we consider general target laws. Historically the emphasis in solving the embedding problem has been in two different directions. We will give a brief description here of some of the work carried out in these two directions; a more detailed survey of the literature can be found in the survey paper, Obłój (2004*b*).

1.2.1 Optimal Embeddings

In addition to the embedding of Skorokhod (1965) there are many examples of embeddings for centred target distributions where the process remains UI. Consequently an interesting question is which embeddings have further maximal or minimal properties. We now list some of these embeddings and their properties:

- Root (1969): based on the construction of barriers, this stopping time minimises $\mathbb{E}T^2$ but is difficult to apply to concrete examples.

- Chacon and Walsh (1976): a construction which generalises Dubins (1968) using techniques from potential theory; we shall look at this construction in more detail in Chapter 4.

- Azéma and Yor (1979*a*): much studied, this embedding has the property that it maximises the law of the maximum among the class of UI embeddings; again we shall return to the study of this process in Chapters 3 and 4.

- Vallois (1983): based on the local time, two similar constructions maximise and minimise the distribution of the local time at zero.

- Perkins (1986): an embedding that has the surprising property that it simultaneously maximises the distribution of the minimum, and minimises the distribution of the maximum; this embedding will form the basis of the work in Chapter 2.

As well as the above embeddings there are many further constructions which build on these (Jacka, 1988; Bass, 1983; Roynette et al., 2002; Hobson, 1998*a*; Brown et al., 2001*a*) or provide different approaches to the same embeddings (Azéma and Yor, 1979*b*; Pierre, 1980; Rogers, 1981; Meilijson, 1983).

1.2.2 Process Generalisations

Another natural direction in which to extend the questions is to consider the same problem for more general processes. Consider the problem of embedding a distribution on some space E into a Markov process on the same space. Immediately this introduces a new complication: consider the problem of embedding a point mass at some non-zero point of $\mathbb{R}^2$ into a Brownian motion in $\mathbb{R}^2$; the process will almost surely avoid the point and we cannot find a finite stopping time which embeds. So in the more general case we must first ask the question of when can we embed. This was the question asked by Rost (1971) who used potential theoretic results to show that there exists a stopping time embedding μ in the process $(X_t)_{t\geq 0}$ with $X_0 \sim \mu_0$ if and only if

$$U\mu \leq U\mu_0,$$

where U is the potential kernel of X. In this full generality very few explicit embeddings exist, although Bertoin and Le Jan (1992) show that when X is a Hunt process[1] starting from a regular, recurrent point, we can construct an embedding which minimises $\mathbb{E}(A_T)$

[1] A Hunt process is a standard process which is quasi-left-continuous on $[0, \infty)$; a special case is a Lévy process.

for any positive continuous additive functional. More recently Obłój and Yor (2004) have constructed an embedding — based on the local time — of a functional of the process, where the process can be from a general class of processes which include for example the Azema martingale.

Further work has considered n-dimensional Brownian motion (Heath, 1974), and several authors have considered (time-homogeneous) diffusions (Grandits and Falkner, 2000; Pedersen and Peskir, 2001; Hambly et al., 2003); as we shall see, this is an example which has close connections with the Brownian case through the technique of scale-change, and in particular to the case where the target distribution is not centred, and possibly not integrable. We shall study this technique in more detain in Section 2.3.

1.3 Applications

1.3.1 Donsker's Invariance Principle

One of the key applications of solutions to the Skorokhod embedding problem is to prove the following theorem:

Theorem 1.4 (Donsker's Invariance Principle)**.** *Let S_n be a simple random walk with a centred step distribution μ of variance 1. Define*

$$S_t^{(n)} = n^{1/2}\left[\left(t-\frac{k}{n}\right)S_{k+1}+\left(\frac{k+1}{n}-t\right)S_k\right], \qquad \frac{k}{n}\leq t\leq\frac{k+1}{n}.$$

Then the processes $(S_t^{(n)})_{0\leq t\leq 1}$ converge weakly to Brownian motion $(B_t)_{0\leq t\leq 1}$ as $n\to\infty$ on the space $C[0,1]$ of continuous functions on $[0,1]$.

A proof of this result is to construct a sequence $0\leq T_1\leq T_2\leq\ldots$ of stopping times such that T_1 is an embedding of μ in a Brownian motion $\tilde{B}_t$ with $\mathbb{E}T_1=1$, T_2-T_1 is an embedding of μ in $(\tilde{B}_{T_1+t}-\tilde{B}_{T_1})$ with $\mathbb{E}T_2=2$, and so on. Since $\mathbb{E}T_n=n$ (and this is crucial) in the limit it is possible to show we have the desired convergence.

Further work in this direction includes a result on the speed of the above convergence (Strassen, 1967), and showing that in fact any local semimartingale is a time change of a Brownian motion, and that these are the only possible time changes (Monroe, 1978).

1.3.2 Optimal Stopping Theory

Let ϕ be a increasing, continuous function and c a continuous function and define $\overline{B}_t = \sup_{s\leq t} B_t$. Then we look to maximise

$$V_T = \mathbb{E}\left(\phi(\overline{B}_T) - \int_0^T c(B_s)\,ds\right) \tag{1.2}$$

over all stopping times for which

$$\mathbb{E}\left(\phi(\overline{B}_T) + \int_0^T c(B_s)\,ds\right) < \infty.$$

This problem was solved originally by Dubins and Schwarz (1988) when $\phi(x) = x$ and $c(x) = c > 0$; their solution is based on the embedding of Azéma and Yor (1979*a*). The solutions when more general functions are considered have been investigated in Peskir (1998, 1999); Meilijson (2003); Obłój (2004*a*), where the solutions are shown (when they are unique) to be the Azema-Yor stopping time for target distributions dependent on the functions ϕ and c. Also examined by some of these authors is the question: for a given μ, what pairs of functions ϕ, c have a solution to (1.2) where the optimal T is an embedding of μ? Again the Azema-Yor solution is of importance in answering this question.

1.3.3 Finance

In mathematical finance asset prices are commonly modelled as a stochastic process and lack of arbitrage conditions suggest that the (discounted) underlying process is a martingale under what is known as the risk-neutral measure. Commonly traded products in the financial markets are European calls, which are contracts based on an underlying asset $(S_t)_{t\geq 0}$. A European call is a contract that entitles the holder to buy the underlying asset at a price K at some fixed future time T_0. Mathematically the payout of such a contract at time T_0 can then be written $(S_{T_0} - K)^+$. We can then price this contract (assuming zero interest rates) as

$$C(K, T_0) = \mathbb{E}^{\mathbb{Q}}(S_{T_0} - K)^+.$$

When the behaviour of the asset is known (under $\mathbb{Q}$) we can calculate $C(K, T_0)$.

In a similar manner, more complicated derivatives can be priced, an example of this being the lookback option which will pay $\sup_{t\leq T_0} S_t$ at time T_0. When the model of the underlying is unknown, but can be assumed to be a continuous martingale (under the risk-neutral measure), Skorokhod embeddings give a method for pricing the lookback option, under the assumption that the call prices $C(K, T_0)$ are known for all strike prices. The idea is that there is some time change under which the underlying is a Brownian motion, and the time T_0 is transformed to some stopping time T of the Brownian motion. The distribution of the process at this time is determined by the call prices. The problem of finding suitable processes is related to finding a stopping time embedding the determined distribution. If, for example, we can find the maximal distribution of the maximum, then this will provide an upper bound on the price of a lookback option. This is precisely the technique used in Hobson (1998*b*) and subsequently extended in Brown et al. (2001*b*) (where knowledge of strike prices at an intermediate time is supposed) and Hobson and Pedersen (2002), which supposes the stock starts in some given distribution. In general upper and lower bounds can be given, together with hedging strategies to exploit any prices outside these ranges.

Further to these examples, Madan and Yor (2002) supposes that the call prices $C(K, t)$ are known both for all strikes and all times thus determining the marginal distributions of the underlying processes. They then show that solutions to the Skorokhod embedding problem can be used to construct (non-unique) processes which meet these marginals.

A further connection to finance can be seen in Cox and Hobson (2004*a*), where assets with a pricing bubble are considered. As well as using the above ideas in pricing options where the underlying process has a financial bubble, it turns out that a key idea in establishing tradable portfolios is to demand (under the risk-neutral measure)

$$x\mathbb{Q}(\inf_{t \leq T} V_t \leq -x) \to 0$$

as $x \to \infty$, where V_t is the value of the portfolio at time t and T is the terminal date of the economy. We will later see that this condition has related interpretations in the theory of embeddings.

1.4 An Overview of the Subsequent Material

The material in this thesis is substantially concerned with embedding in Brownian motion and the following question:

Key Question. Let μ be any probability measure on $\mathbb{R}$. Can we find a general class of embeddings which includes 'nice' stopping times such as Example 1.2 but excludes 'nasty' examples like Example 1.1?

1.4.1 Chapter 2: The Minimax-Maximin Solution

We begin by examining an embedding of Perkins (1986). He shows that, given a centred, integrable target distribution μ we are able to construct functions γ_+, γ_- such that the stopping time

$$T_P = \inf\{t > 0 : B_t \notin (-\gamma_+(\overline{B}_t), \gamma_-(-\underline{B}_t))\}$$

is an embedding. Here we have defined the supremum and infimum processes of B:

$$\begin{aligned} \overline{B}_t &= \sup_{s \le t} B_t; \\ \underline{B}_t &= \inf_{s \le t} B_t. \end{aligned}$$

The remarkable property of this embedding is that it simultaneously minimises

$$\mathbb{P}(\overline{B}_T \ge x)$$

and maximises

$$\mathbb{P}(\underline{B}_T \ge -x)$$

over all $x \ge 0$ and all embeddings T of μ. We begin by showing that we can extend these results to all (not necessarily integrable) target measures.

We then discuss a related problem: given a regular, time-homogeneous diffusion on an interval $I \subseteq \mathbb{R}$ there exists a *scale function* $s(x)$ which is a function mapping the diffusion to a local martingale, and therefore a time-change of a Brownian motion. We might ask when are we able to construct an embedding of the diffusion for a particular target distribution, and via a scale change this essentially becomes a question of finding an embedding of the new local martingale under certain extra conditions. We show how this can be done using the method of scale change, and how the embedding we use fits nicely with this technique.

Finally we consider the problem of finding H^p-embeddings. An embedding T (or its associated process $X_{t\wedge T}$) is an H^p-embedding if and only if

$$\mathbb{E}(\sup_{t \le T} |X_t|^p) < \infty.$$

The stopping times we introduce minimise the law of $\sup_{t\leq T}|X_t|^p$, and therefore if there exists an H^p-embedding for the distribution μ, the minimax solution will be such a solution. This is a question that has been considered by Perkins (1986) for the centred, Brownian case, who shows that for $p>1$ the embedding is in H^p if and only if $\mu \in \mathcal{L}^p$; the case where $p=1$ is more subtle and the embedding is in H^p if and only if

$$\int_0^\infty y^{-1}\left|\int_{-\infty}^\infty x\mathbf{1}_{\{|x|\geq y\}}\,\mu(dx)\right| < \infty.$$

We consider the problem of the existence of H^p-embeddings for diffusions and give necessary and sufficient conditions on the diffusion and the target law for the existence of H^p-embeddings. Under certain additional assumptions we are able to give simple and often equivalent conditions.

1.4.2 Chapter 3: Minimality and Azema-Yor Type Embeddings

In this chapter we are again interested in the problem of embedding non-centred target distributions. However in this section we consider a slightly different optimality condition. Instead of trying to minimise the distribution of the maximum we now consider the problem of maximising this distribution. Of course in the general case this problem is degenerate: we can ensure that

$$\mathbb{P}(\overline{B}_T \geq x) = 1$$

for any x simply by considering stopping times of the form 'wait until the process hits x, wait until it returns to 0 and then use any desired embedding.' Clearly we cannot attain equality for all x since $T<\infty$ a.s..

In the Brownian case, with a centred target distribution, this issue is resolved by considering only stopping times for which $B_{t\wedge T}$ is UI. Then the solution to the problem is given by the embedding of Azéma and Yor (1979*a*,*b*). This is the embedding constructed using the *barycentre* function:

$$\Psi(x) = \begin{cases} \frac{1}{\mu([x,\infty))}\int_{\{u\geq x\}} u\,\mu(du) & \mu([x,\infty))>0 \\ x & \mu([x,\infty))=0. \end{cases}$$

Then the Azema-Yor stopping time is defined to be:

$$T_{AY} := \inf\{t>0 : \overline{B}_t \geq \Psi(B_t)\},$$

which maximises $\mathbb{P}(\overline{B}_T \geq x)$ for all $x \geq 0$ over the class of embeddings which are UI.

When we consider the problem of embeddings with non-centred target distributions, in order to make the question sensible we need to find a natural class of embeddings which rules out pathological examples. One constraint suggested in Pedersen and Peskir (2001) is to consider only stopping times for which $\mathbb{E}(\overline{B}_T) < \infty$; however this seems unnatural in this context — we could be ruling out stopping times simply because they are too good! Instead we suggest the following concept originally proposed by Monroe (1972).

Definition 1.5. A stopping time T for the process X is *minimal* if whenever $S \leq T$ is a stopping time such that X_S and X_T have the same distribution then $S = T$ a.s..

One of the key results concerning minimal embeddings in Monroe (1972) is that — in the case where the target distribution is centred — an embedding is minimal if and only if it is UI. We show that minimal embeddings in the more general situation of non-centred (but still integrable) target distributions have the properties we desire: that is, we can find an extension to the Azema-Yor stopping time which is itself minimal and maximises the maximum among the class of minimal stopping times. Further we show similarly that the embedding of Jacka (1988) can be extended to allow the construction of stopping times that maximise the distribution of $\sup_{t \leq T} h(B_t)$ for any function h.

We also consider again the implications of minimal embeddings for diffusions. If we consider the embedding to be a scale change of a Brownian motion, when the diffusion is transient and solutions exist there is a one-to-one correspondence between embeddings of the diffusions and minimal embeddings in the Brownian scale. The techniques established in the Brownian case are then easily extended to the diffusion case — one of the nice properties of minimality being that it is unaltered by scale change techniques.

Also included in this chapter is a partially constructive proof that minimal stopping times exist; this compares with a non-constructive proof due to Monroe (1972).

1.4.3 Chapter 4: Generalised Starting Distributions and the Chacon-Walsh Construction

Finally we consider a further generalisation of the problem: we suppose in this section that the Brownian motion starts in some distribution μ_0. Then we can ask when is an embedding minimal. The conditions given previously are no longer sufficient, as can be seen by the following example. Consider starting with mass divided equally between -1 and 1, and with stopping distribution consisting of a unit mass at 0; both distributions are centred, and it is clear that the only minimal embedding is to stop the first time the process hits 0, yet this is not a UI stopping time. Clearly we have to refine the conditions of Chapter 3.

It will turn out that necessary and sufficient conditions to be minimal in this setting are closely related to the 'potential' of the measures. For the purposes of this work we define the potential of a measure to be

$$u_\mu(x) = -\int |y - x| \, \mu(dx).$$

Using these functions we are able to specify intervals on which a minimal process will remain, and its behavior on these intervals is also specified.

The potential functions are known to play an important role in the embedding of centred target distributions, and Chacon and Walsh (1976) provides a construction of an embedding in the case where $u_{\mu_0}(x) \geq u_\mu(x)$ for all $x \in \mathbb{R}$. We extend their construction to allow it to be used in the general case, using our results on minimality to give a simple way of determining whether a stopping time constructed using these techniques is indeed minimal.

Finally we show that by taking limits in the new construction some of the classical embeddings, such as the Azema-Yor embedding and the Vallois embedding, can be constructed. This has several advantages: it allows extensions of the constructions to be generated easily; it allows us to deduce that the embeddings are minimal; for some embeddings, such as the Azema-Yor embedding, we are able to prove optimality properties among the class of minimal embeddings.

Chapter 2

The Minimax-Maximin Solution and Applications to Diffusions

(This work has appeared in Cox and Hobson (2004b))

In this Chapter we extend the embedding first described in Perkins (1986) to the case where the target distribution is not centred or even integrable. We are also able to show that the embedding inherits the optimality properties of the original embedding.

In this context it is natural to consider the problem of embedding in time-homogeneous diffusions. Via the method of scale change this can be related to the Brownian case, and when the extended Perkins embedding is used for the Brownian scale-change, the diffusion inherits the optimality properties of the Brownian case. This allows us to address further questions of interest, including the existence of H^p-embeddings for a given diffusion and target distribution.

2.1 Introduction

We begin this chapter by considering the embedding problem for a continuous local martingale $(M_t)_{t\geq 0}$ on $\mathbb{R}$, vanishing at 0. When the quadratic variation $\langle M\rangle_\infty = \infty$ a.s. we show that we are able to extend the embedding of Perkins from the case where the target distribution is centred to the general case where there are no restrictions on the distribution. We are also able to show that the embedding is optimal in the sense that for any embedding T and $x \geq 0$

$$\begin{aligned}\mathbb{P}(\overline{M}_T \geq x) &\geq \mathbb{P}(\overline{M}_{T_P} \geq x);\\ \mathbb{P}(\underline{M}_T \geq -x) &\leq \mathbb{P}(\underline{M}_{T_P} \geq -x),\end{aligned}$$

where T_P is the (extended) Perkins embedding and

$$\overline{M}_t = \sup_{s\leq t} M_s; \tag{2.1}$$

$$\underline{M}_t = \inf_{s\leq t} M_s. \tag{2.2}$$

This property has the additional consequence that the stopping time T_P is minimal (see Definition 1.5 and further discussion in Chapters 3 and 4), since any other embedding must be larger than T_P on some set of positive probability.

The second purpose of this chapter is to consider the embedding of μ in a one-dimensional diffusion. The main technique is to use a change of scale to reduce the problem to the local-martingale case, and under this transformation it is completely natural for the target measure to have non-zero mean in the local-martingale (or Brownian) scale. We will see that our embedding is a natural one to use in this situation, and we are able to identify the cases where it is possible to embed a given target distribution, thus rederiving a result in Pedersen and Peskir (2001). We also identify some properties of the maximum and minimum of the processes in these cases. Our results in this direction can be seen as an extension of the results in Grandits and Falkner (2000) (for drifting Brownian motion) and Pedersen and Peskir (2001). In this last paper the authors use an extension of the Azema-Yor embedding which may not be defined in certain cases of interest. Thus our construction of a Skorokhod embedding is both different to, and more general than, the embedding in Pedersen and Peskir (2001).

Finally we use the optimal properties of the embedding, together with the scale change techniques to deduce when it is possible to construct a H^p-embedding, i.e. given a diffusion process Y and a target law ν when does there exists a stopping time T such that $Y_T \sim \nu$ and $\mathbb{E}\sup_t |Y_{t\wedge T}|^p < \infty$. While the construction gives us explicit formulae for the distribution of the maxima and minima, these are often difficult to calculate, and in some cases we are able to give conditions which are simpler to verify.

2.2 Embedding a General Target Measure in Brownian Motion

Consider first the problem of embedding a target distribution μ in a one-dimensional local martingale $(M_t)_{t\geq 0}$, $M_0 = 0$ a.s.. We make no assumptions on μ other than that $\mu(\mathbb{R}) = 1$, and that μ has no atom at 0. In fact this second assumption can be avoided by stopping immediately according to some independent randomisation with suitable probability, and then using the construction to embed the remaining mass of μ, conditional on not stopping at 0. Clearly such a construction is necessary in any stopping time that will minimise the maximum, and maximise the minimum.

For a general local martingale the above conditions are not sufficient to ensure that an embedding exists. However a sufficient condition for the existence of an embedding for any μ is that our local martingale almost surely has infinite quadratic variation. Since any local martingale is simply a time change of Brownian motion, this just ensures that our time change does not stall.

We begin by defining a series of functions. Let

$$\kappa(x) = \begin{cases} \int_{\{u\geq 0\}} (x \wedge u)\,\mu(du) & : \quad x \geq 0; \\ \int_{\{u<0\}} (|x| \wedge |u|)\,\mu(du) & : \quad x < 0. \end{cases} \tag{2.3}$$

Then $\kappa(x)$ is increasing and concave on $\{x \geq 0\}$, decreasing and concave on $\{x \leq 0\}$ and continuous on $\mathbb{R}$ (see Figures 2-1 and 2-2). It is also differentiable Lebesgue-almost-everywhere and:

$$\kappa'(x)_+ = \begin{cases} \mu((x,\infty)) & : \quad x \geq 0; \\ -\mu((-\infty,x]) & : \quad x < 0; \end{cases} \tag{2.4}$$

$$\kappa'(x)_- = \begin{cases} \mu([x,\infty)) & : \quad x > 0; \\ -\mu((-\infty,x)) & : \quad x \leq 0, \end{cases} \tag{2.5}$$

where $\kappa'(x)_-, \kappa'(x)_+$ are the left and right derivatives respectively. In particular, the points at which $\kappa(x)$ is not differentiable are precisely the atoms of out target distribution. We also note that $\kappa(\infty) < \infty$ if and only if our target distribution satisfies $\int_{\{x\geq 0\}} x\,\mu(dx) < \infty$, when $\kappa(\infty) = \int_{\{x\geq 0\}} x\,\mu(dx)$, and similarly $\kappa(-\infty) = \int_{u<0} |u|\mu(du)$ when this is finite. Finally, we have $\kappa(\infty) = \kappa(-\infty) < \infty$ if and only if $\mu \in \mathcal{L}^1$ and μ is centred.

For $\lambda > 0$, define the following quantities:

$$\gamma_+(\lambda) = \operatorname*{argmin}_{x>0} \left\{ \frac{\kappa(\lambda) - \kappa(-x)}{\lambda - (-x)} \right\}, \tag{2.6}$$

$$\gamma_-(\lambda) = \operatorname*{argmax}_{x>0} \left\{ \frac{\kappa(x) - \kappa(-\lambda)}{x - (-\lambda)} \right\}, \tag{2.7}$$

$$\theta_+(\lambda) = -\inf_{x>0} \left\{ \frac{\kappa(\lambda) - \kappa(-x)}{\lambda - (-x)} \right\}, \tag{2.8}$$

$$\theta_-(\lambda) = \sup_{x>0} \left\{ \frac{\kappa(x) - \kappa(-\lambda)}{x - (-\lambda)} \right\}, \tag{2.9}$$

$$\begin{aligned} \mu_+(\lambda) &= \theta_+(\lambda) + \mu([\lambda,\infty)), \\ &= -\frac{\kappa(\lambda) - \kappa(-\gamma_+(\lambda))}{\lambda - (-\gamma_+(\lambda))} + \kappa'(\lambda)_-, \end{aligned} \tag{2.10}$$

$$\begin{aligned} \mu_-(\lambda) &= \mu((-\infty,-\lambda]) + \theta_-(\lambda), \\ &= -\kappa'(-\lambda)_+ + \frac{\kappa(\gamma_-(\lambda)) - \kappa(-\lambda)}{\gamma_-(\lambda) - (-\lambda)}, \end{aligned} \tag{2.11}$$

where $\operatorname{argmin}_x f(x)$ is the value of x which minimises the function f and $\operatorname{argmax}_x f(x)$ is the value which maximises the function f. If the minimising (respectively maximising) x in (2.6) (resp. (2.7)) is not unique then we take the smallest such x. If there is no minimising x, then the function we are minimising is decreasing (resp. increasing) as $x \to \infty$, and we define $\gamma_+(\lambda) = \infty$ (resp. $\gamma_-(\lambda) = \infty$). In this case we also define $\theta_+(\lambda) = 0$ (resp. $\theta_-(\lambda) = 0$).

Remark 2.1. Although we have given formal definitions these quantities are best described pictorially. Given $\lambda > 0$, we consider points $(y, \kappa(y))$ for $y < 0$ and more specifically the line segment joining $(y, \kappa(y))$ with $(\lambda, \kappa(\lambda))$. As y ranges over the negative reals we let $\theta_+(\lambda)$ be the steepest possible downward slope of this line segment, and we let $\gamma_+(\lambda)$ be the absolute value of the x-coordinate of the point where this maximum is attained. See Figures 2-1 and 2-2.

The quantities $\theta_-(\lambda)$ and $\gamma_-(\lambda)$ are obtained by reflecting the picture. Alternatively, if we define the measure $\tilde{\mu}((-\infty, x]) = \mu([-x,\infty))$ then we obtain a correspondence between the pairs of definitions above — that is $\gamma_-^{\mu}(\lambda) = \gamma_+^{\tilde{\mu}}(\lambda)$, $\theta_-^{\mu}(\lambda) = \theta_+^{\tilde{\mu}}(\lambda)$ and $\mu_-(\lambda) = \tilde{\mu}_+(\lambda)$, with the obvious extension of the notation.

Remark 2.2. It is only possible to have $\gamma_+(\lambda) = \infty$ when μ satisfies $\int_{\{x\geq 0\}} x\,\mu(dx) >$

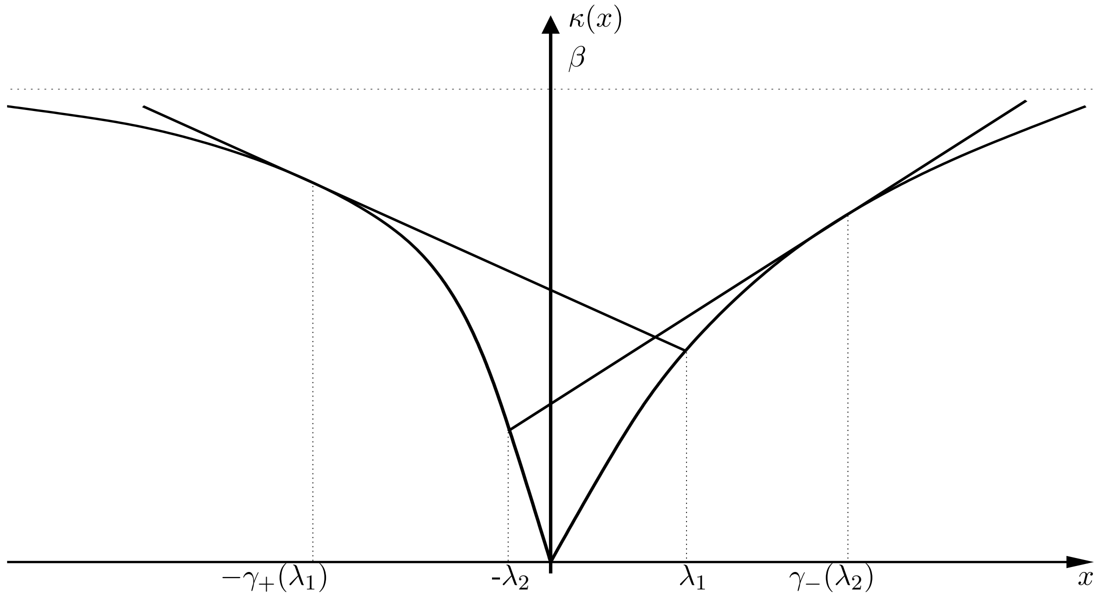

Figure 2-1: $\kappa(x)$ for a centred non-atomic measure. As $|x| \to \infty$, $\kappa(x)$ is asymptotic to β, where $\beta = \int_{\{x \geq 0\}} x\,\mu(dx)$.

$\int_{\{x \leq 0\}} |x|\,\mu(dx)$, as in Figure 2-2. If this is true, then $\gamma_+(\lambda) = \infty$ for all λ such that $\kappa(\lambda) > \int_{\{x \leq 0\}} |x|\,\mu(dx)$ (and if the support of μ is not bounded below, also when equality holds).

We take this opportunity to record some further relationships between the various quantities defined in (2.6) to (2.11). It follows from (2.6) and (2.7) that for $\lambda > 0$:

$$-\kappa'(-\gamma_+(\lambda))_- \leq \theta_+(\lambda) \leq -\kappa'(-\gamma_+(\lambda))_+, \tag{2.12}$$

$$\kappa'(\gamma_-(\lambda))_+ \leq \theta_-(\lambda) \leq \kappa'(\gamma_-(\lambda))_-, \tag{2.13}$$

so there is equality in (2.12) or (2.13) when there is no atom of μ at $-\gamma_+(\lambda)$ or $\gamma_-(\lambda)$ respectively. From Figure 2-2 it is clear that if there is an atom of μ at $-\gamma_+(\lambda)$ then κ has a kink there, and $-\theta_+(\lambda)$ is then the gradient of the line joining $(-\gamma_+(\lambda), \kappa(-\gamma_+(\lambda)))$ and $(\lambda, \kappa(\lambda))$. Further, for $\lambda > 0$ such that $\gamma_+(\lambda), \gamma_-(\lambda) < \infty$, we have

$$\kappa(\lambda) = \kappa(-\gamma_+(\lambda)) - (\lambda + \gamma_+(\lambda))\theta_+(\lambda), \tag{2.14}$$

$$\kappa(-\lambda) = \kappa(\gamma_-(\lambda)) - (\lambda + \gamma_-(\lambda))\theta_-(\lambda). \tag{2.15}$$

Note that as a simple consequence of these equalities, $\kappa(\lambda) \leq \kappa(-\gamma_+(\lambda))$ and $\kappa(-\lambda) \leq \kappa(\gamma_-(\lambda))$.

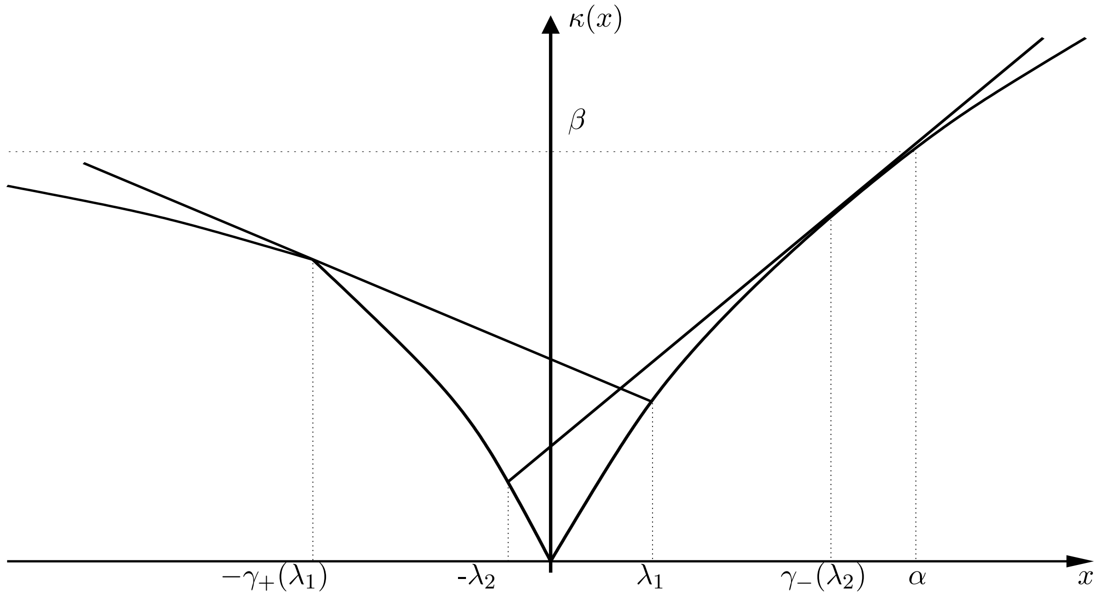

Figure 2-2: $\kappa(x)$ for a non-integrable measure with an atom at $-\gamma_+(\lambda_1)$. As $x \to \infty$, $\kappa(x) \to \int_{\{x \geq 0\}} x\,\mu(dx) = \infty$, while as $x \to -\infty$, $\kappa(x)$ is asymptotic to the level $\beta = -\int_{\{x \leq 0\}} x\,\mu(dx)$, which for this example is taken to be finite. The point α is such that $\kappa(\alpha) = \beta$, and for all $\lambda > \alpha$, $\gamma_+(\lambda) = \infty$.

Remark 2.3. By considering Figures 2-1 and 2-2, we see that alternative definitions for $\gamma_+(\lambda)$, $\gamma_-(\lambda)$, $\theta_+(\lambda)$ and $\theta_-(\lambda)$ are

$$\gamma_+(\lambda) = -\sup\left\{x < 0 : \frac{\kappa(\lambda) - \kappa(x)}{\lambda - x} \leq \kappa'(x)_+\right\}, \tag{2.16}$$

$$\gamma_-(\lambda) = \inf\left\{x > 0 : \frac{\kappa(x) - \kappa(-\lambda)}{x - (-\lambda)} \geq \kappa'(x)_-\right\}, \tag{2.17}$$

$$\theta_+(\lambda) = -\frac{\kappa(\lambda) - \kappa(-\gamma_+(\lambda))}{\lambda - (-\gamma_+(\lambda))}, \tag{2.18}$$

$$\theta_-(\lambda) = \frac{\kappa(\gamma_-(\lambda)) - \kappa(-\lambda)}{\gamma_-(\lambda) - (-\lambda)}. \tag{2.19}$$

As a result it is easy to see that, in the case where μ is centred, these quantities are identical to the quantities defined in Perkins (1986), where the quantity $q_+(\lambda)$ defined in Perkins (1986) satisfies $\theta_+(\lambda) = q_+(\lambda) + \mu((-\infty, -\gamma_+(\lambda)))$.

Our first theorem shows that for any target measure μ there is an embedding which simultaneously stochastically maximises the distribution of the minimum, and minimises the distribution of the maximum.

Theorem 2.4. *(i) Let $(M_t)_{t\geq 0}$ be a continuous local martingale, vanishing at zero, and let T be a stopping time such that $M_T \sim \mu$. Then, for all $\lambda \geq 0$, the following hold:*

$$\mathbb{P}(\overline{M}_T \geq \lambda) \geq \mu_+(\lambda) \tag{2.20}$$
$$\mathbb{P}(-\underline{M}_T \geq \lambda) \geq \mu_-(\lambda) \tag{2.21}$$

(ii) For a continuous local martingale, M_t, vanishing at zero and such that $\langle M\rangle_\infty = \infty$ a. s., define the stopping time

$$T = \inf\{t > 0 : M_t \notin (-\gamma_+(\overline{M}_t), \gamma_-(-\underline{M}_t))\}. \tag{2.22}$$

Then the stopped process M_T has distribution μ, and equality holds in (2.20) *and* (2.21)*.*

Remark 2.5. When μ is centred, the fact that the quantities γ_+ and γ_- agree with those in Perkins (1986), and the fact that in this case T as defined in (2.22) is the Perkins stopping time, means that we know that T embeds μ. Moreover we know that T minimises the law of the maximum, and maximises the law of the minimum. These results follow directly from Theorems 3.7 and 3.8 in Perkins (1986). The content of Theorem 2.4 is that these results can be extended to any choice of μ.

Remark 2.6. We may think of $\theta_+(\lambda)$ and $\theta_-(\lambda)$ as probabilities, and in particular, for the embedding defined in (2.22), $\theta_+(\lambda)$ is the probability that our process stops below $-\gamma_+(\lambda)$ but with a maximum above λ. If μ has no atom at $-\gamma_+(\lambda)$ then for this construction the maximum will be above λ if and only if our final value is above λ or below $-\gamma_+(\lambda)$. However if there is an atom at $-\gamma_+(\lambda)$, the process may stop there without previously having reached λ. This event is represented graphically by the fact that there are multiple tangents to κ at $-\gamma_+(\lambda)$. Also, when $\gamma_+(\lambda) = \infty$ for some λ, if the supremum of our process gets above λ before stopping then our stopping rule becomes simply to wait until we reach some upper level, dependent on the infimum.

An alternative way to visualise the stopping time in (2.22) is shown in Figure 2-3. We think of the process $(-\underline{M}_t, \overline{M}_t)$, and define the stopping time to be the first time it leaves the region defined via γ_+ and γ_- as shown.

The first half of the proof of Theorem 2.4 is a consequence of the following lemma.

Lemma 2.7. *Let $(M_t)_{t\geq 0}$ be a continuous local martingale. Suppose that M vanishes at zero, M converges a.s., and that $M_\infty \sim \mu$, for some probability measure μ on $\mathbb{R}$.*

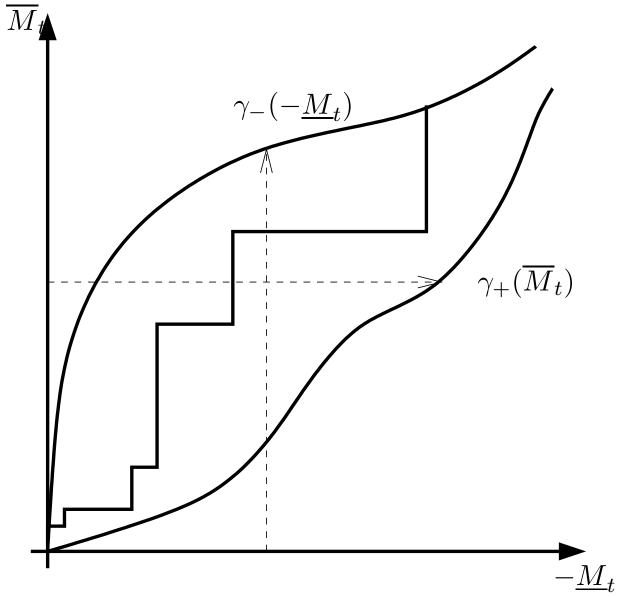

Figure 2-3: The path of the process in the $(-\underline{M}_t, \overline{M}_t)$-space. T is the first time this process leaves the region.

Then, for $\lambda > 0$,

$$\mathbb{P}(\overline{M}_\infty \geq \lambda) \geq \mu_+(\lambda), \tag{2.23}$$

$$\mathbb{P}(-\underline{M}_\infty \geq \lambda) \geq \mu_-(\lambda), \tag{2.24}$$

where $\overline{M}_\infty = \sup_s M_s$, and $\underline{M}_\infty = \inf_s M_s$.

Proof. For $x < 0 < \lambda$, we define $H_\lambda = \inf\{t > 0 : M_t = \lambda\}$ where we take $\inf \emptyset = \infty$. By examining on a case by case basis, we find that the following inequality holds:

$$\mathbf{1}_{\{\overline{M}_\infty \geq \lambda\}} \geq \mathbf{1}_{\{M_\infty \geq \lambda\}} + \frac{1}{\lambda - x}\big[M_{H_\lambda} - (\lambda \wedge M_\infty)\mathbf{1}_{\{M_\infty \geq 0\}} + (|M_\infty| \wedge |x|)\mathbf{1}_{\{M_\infty < 0\}}\big].$$

After taking expectations, this implies that

$$\mathbb{P}(\overline{M}_\infty \geq \lambda) \geq \kappa'(\lambda)_- + \frac{1}{\lambda - x}\mathbb{E}M_{H_\lambda} - \frac{\kappa(\lambda) - \kappa(x)}{\lambda - x}.$$

Now $M_{t \wedge H_\lambda}$ is a local martingale bounded above, and hence a submartingale, so $\mathbb{E}M_{H_\lambda} \geq M_0 = 0$. Substituting this in the above equation, we get:

$$\mathbb{P}(\overline{M}_\infty \geq \lambda) \geq \kappa'(\lambda)_- - \frac{\kappa(\lambda) - \kappa(x)}{\lambda - x},$$

and since x is arbitrary,

$$\begin{aligned}\mathbb{P}(\overline{M}_\infty \geq \lambda) &\geq \kappa'(\lambda)_- + \sup_{x<0}\left\{\frac{\kappa(x)-\kappa(\lambda)}{\lambda - x}\right\}\\ &\geq \mu([\lambda,\infty)) + \theta_+(\lambda) = \mu_+(\lambda),\end{aligned}$$

which is (2.23).

We may deduce (2.24) using the correspondence $\mu \mapsto \tilde{\mu}$. □

Remark 2.8. In particular, for equality to hold for fixed λ in the above, we must have

(i) if $\overline{M}_\infty \geq \lambda$, either $M_\infty \geq \lambda$ or $M_\infty \leq -\gamma_+(\lambda)$ a.s.,

(ii) if $\overline{M}_\infty < \lambda$, $M_\infty \geq -\gamma_+(\lambda)$ a.s.,

(iii) $\mathbb{E}M_{H_\lambda} = 0$, so that $M_{t\wedge H_\lambda}$ is a true martingale.

It can be seen that these will hold simultaneously for all λ in the case where the stopping time is that given in Theorem 2.4, and that this is almost surely the only stopping time where (2.23) and (2.24) hold.

Proof of Theorem 2.4. We apply Lemma 2.7 to the process $(M_{T\wedge t})_{t\geq 0}$, which allows us to deduce (2.20) and (2.21).

For the second part of the theorem recall that if μ is centred then the Theorem follows from Theorems 3.7 and 3.8 in Perkins (1986). In the case when μ is not centred define

$$\begin{aligned}\xi_+^n &= \inf\left\{x : \mu([x,\infty)) \leq \frac{1}{2n}\right\},\\ \xi_-^n &= \sup\left\{x : \mu((-\infty,x]) \leq \frac{1}{2n}\right\},\end{aligned}$$

and, for n sufficiently large, consider a sequence of measures μ^n satisfying:

(i) $\mu^n((\alpha,\beta)) = \mu((\alpha,\beta)),\ \xi_-^n < \alpha \leq \beta < \xi_+^n$;

(ii) $\mu^n([\xi_-^n,\xi_+^n]) = \mu^n([(-n)\wedge\xi_-^n, n\vee\xi_+^n]) = \frac{n-1}{n}$;

(iii) $\mu^n(\{\xi_\pm^n\}) \leq \mu(\{\xi_\pm^n\})$;

(iv) $\int x\,\mu^n(dx) = 0$;

(v) $\int |x|\,\mu^n(dx) < \infty$.

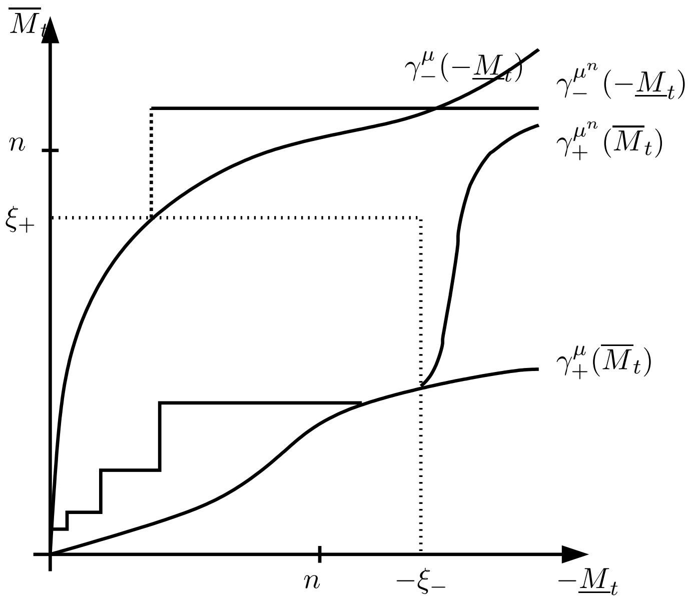

Figure 2-4: The path of the process in the $(-\underline{M}_t, \overline{M}_t)$-space, showing boundaries to embed both μ and μ^n. We have shown here a possible choice of μ^n in the case where $\xi_+ < n < (-\xi_-)$.

We can construct such a sequence by redistributing the mass that lies in the tails of μ as follows: each μ^n agrees with μ on the interval (ξ_-^n, ξ_+^n), and mass is placed at the endpoints of this interval to satisfy *(ii)* and *(iii)* if there are atoms here; the remaining mass is then placed outside the interval $[(-n) \wedge \xi_-^n, n \vee \xi_+^n]$ in such a way as to ensure that *(iv)* and *(v)* hold.

For the rest of this section a superscript n will denote the fact that a quantity is calculated relative to the measure μ^n.

Note that if we can construct μ^n in such a way that $\mu^n(\mathbb{R}_-) = \mu(\mathbb{R}_-)$ then we find that $\kappa^n(x) \equiv \kappa(x)$ on $[\xi_-^n, \xi_+^n]$. However it is not possible to construct μ^n with this additional property if $\mu(\mathbb{R}_-) = 0$ or 1, and in that case we need a more general argument.

Suppose $\mu^n(\mathbb{R}_-) - \mu(\mathbb{R}_-) = \psi_n$ for some number $\psi_n \in (-1/2n, 1/2n)$. Then $\kappa^n(x) = \kappa(x) - \psi_n x$ for $x \in [\xi_-^n, \xi_+^n]$. If both λ and $\gamma_+(\lambda)$ lie in this interval then it is clear from (2.6) that $\gamma_+^n(\lambda) = \gamma_+(\lambda)$. Conversely if $\gamma_+(\lambda) = \infty$, then $\gamma_+^n(\lambda) \geq n$. Similar results hold for γ_-^n.

We define the stopping times associated with these measures,

$$T^n := \inf\{t > 0 : M_t \notin (-\gamma_+^n(\overline{M}_t), \gamma_-^n(-\underline{M}_t))\},$$

so that $M_{T^n} \sim \mu^n$. Note that if $M_{T^n} \in [(-n) \wedge \xi_-^n, n \vee \xi_+^n]$, then $T = T^n$ a. s. (see Figure 2-4). However this implies that $\mathbb{P}(T = T^n) \to 1$, since these intervals are increasing to cover the whole of $\mathbb{R}$. Together with the fact that $\mu^n([\lambda, \infty)) \to \mu([\lambda, \infty))$, we conclude that $M_T \sim \mu$.

Finally, we need to show that our process attains equality in (2.20) and (2.21). Fix $\lambda > 0$. We know that

$$\mathbb{P}(\overline{M}_{T^n} \geq \lambda) = \mu_+^n(\lambda) = \mu^n([\lambda, \infty)) + \theta_+^n(\lambda)$$

and since $\mathbb{P}(T^n = T) \geq (n-1)/n$, we have $\mathbb{P}(\overline{M}_{T^n} \geq \lambda) \to \mathbb{P}(\overline{M}_T \geq \lambda)$. Moreover $\mu^n([\lambda, \infty)) \to \mu([\lambda, \infty))$ so that in order to prove

$$\mathbb{P}(\overline{M}_T \geq \lambda) = \mu([\lambda, \infty)) + \theta_+^\mu(\lambda) = \mu_+(\lambda), \tag{2.25}$$

it is sufficient to show that $\theta_+^n(\lambda) \to \theta_+^\mu(\lambda)$ as $n \to \infty$. Now, when $x \in [\xi_-^n, \xi_+^n]$, we have $\kappa^n(x) - \kappa(x) = \psi_n x$ and for x outside this range $(\kappa^n)' - \kappa' \leq 1/n$. Hence

$$|\kappa^n(x) - \kappa(x)| \leq \frac{|x|}{n},$$

for all x. As a corollary, for $x < 0 < \lambda$,

$$\left| \frac{\kappa^n(\lambda) - \kappa^n(x)}{\lambda - x} - \frac{\kappa(\lambda) - \kappa(x)}{\lambda - x} \right| \leq \frac{1}{n},$$

from which it follows that

$$|\theta_+^{\mu^n}(\lambda) - \theta_+^\mu(\lambda)| \leq \frac{1}{n}.$$

using the representation (2.18).

As before we can also show (2.21) holds by using the correspondence $\mu \mapsto \tilde{\mu}$. □

Example 2.9. We demonstrate the new embedding by constructing a stopping time for a non-integrable, non-symmetric distribution. In this case we use a parametrised Cauchy distribution to give positive and negative tails, with different parameters for each tail.

We note that for a scaled Cauchy distribution on the half-line

$$\int_0^\infty \frac{1}{\pi(1 + (ax)^2)}\, dx = \frac{1}{2a}$$

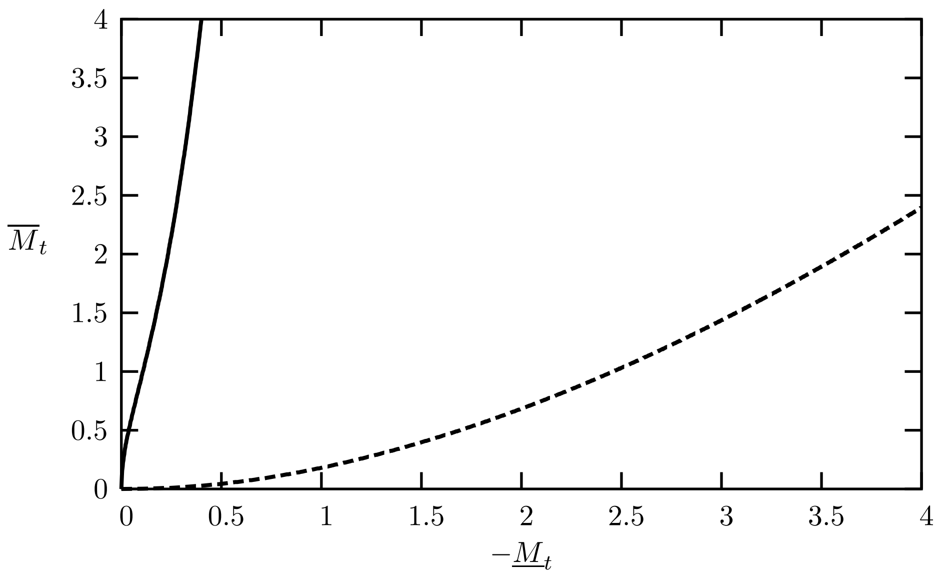

Figure 2-5: $\gamma_-(-\underline{M}_t)$ and $\gamma_+(\overline{M}_t)$ for the distribution f defined in (2.26) with $a = 0.75$ and $b = 1.5$.

and consider a distribution with density

$$f(x) = \begin{cases} \frac{1}{\pi(1+(ax)^2)} & : x \geq 0 \\ \frac{1}{\pi(1+(bx)^2)} & : x < 0 \end{cases} \tag{2.26}$$

where $\frac{1}{2a} + \frac{1}{2b} = 1$. We note that neither tail of the distribution is integrable. We can compute the function κ for this class of densities:

$$\kappa(y) = \begin{cases} \frac{\ln(1+(ay)^2)+(ay)(\pi-2\arctan(ay))}{2\pi a^2} & : y \geq 0; \\ \frac{\ln(1+(by)^2)-(by)(\pi-2\arctan(-by))}{2\pi b^2} & : y < 0. \end{cases}$$

We can use this to find (computationally) the functions γ_- and γ_+. These are shown in Figure 2-5. We observe that the functions move apart very rapidly, in contrast to integrable, centered distributions where the functions would be expected to move closer in the tails; this is partly a consequence of the fact that T is generally 'very' large.

2.3 Applications to Diffusions

We now work with the class of regular (time-homogeneous) diffusions (see Rogers and Williams (2000*b*), V. 45) $(Y_t)_{t\geq 0}$ on an interval $I \subseteq \mathbb{R}$, with absorbing or inaccessible endpoints, and vanishing at zero. The extension to reflecting endpoints is possible, although we leave the extension of the theory to the reader. Consider the problem of determining when and how we may embed a distribution ν on I° in the diffusion. Since the diffusion is regular, there exists a continuous, strictly increasing scale function $s : I \to \mathbb{R}$ such that $M_t = s(Y_t)$ is a diffusion on natural scale on $s(I)$. We may also choose s such that $s(0) = 0$. In particular, M_t is (up to exit from the interior of $s(I)$) a time change of a Brownian motion, with strictly positive speed measure.

If we now define the measure μ on $s(I)$ by

$$\mu(A) = \nu(s^{-1}(A)), \qquad A \subseteq s(I), \text{ Borel},$$

then our problem is equivalent to that of embedding μ in a Brownian motion before it leaves $s(I)^\circ$. This is because M is a local martingale on $s(I)^\circ$, and hence a time change of a Brownian motion on $s(I)^\circ$, and if we construct a stopping time T such that $M_T = s(Y_T) \sim \mu$, then $Y_T \sim \nu$. In this context it makes sense to consider ν and μ as measures on $\mathbb{R}$ which place all their mass on I° and $s(I)^\circ$ respectively. Our approach will be to use the embedding we established in Theorem 2.4 to embed μ in the local martingale M, and our first step will be to transfer the framework of the previous section to our new setting. This framework for embedding in diffusions was first suggested in Azéma and Yor (1979*b*).

An advantage of using the embedding we established in Section 2.2 in this situation is that, because we have a strictly increasing scale function, the properties of the maximum and the minimum are preserved. In particular, this transformed stopping time will maximise the distribution of the minimum, and minimise the distribution of the maximum of the process $(Y_{T\wedge t})$ among all stopping times of Y_t with $Y_T \sim \nu$. It is also important to have an embedding which works when the mean of the target distribution is non-zero, since under the scale change transition described above the properties of the target distribution will be altered — it is perfectly natural for a target distribution not to be centred or even integrable under this transformation.

The first question that it is necessary to ask is: when is it possible to embed a given target law? This is exactly the question considered by Rost (1971) using potentials, but we want a more direct criterion. In the diffusion case it is no longer possible to embed all target laws, as can be witnessed by considering the problem of embedding unit mass at -1 in Brownian motion with positive drift. The result we need was first proved in Pedersen and Peskir (2001).

Lemma 2.10 (Pedersen and Peskir (2001), Theorem 2.1)**.** *There are three different* *cases:*

(i) $s(I)^\circ = \mathbb{R}$*, in which case the diffusion is recurrent, and we can embed any distribution* ν *on* I° *in* Y*.*

(ii) $s(I)^\circ = (-\infty, \alpha)$ *(respectively* (α, ∞)*) for some* $\alpha \in \mathbb{R}$*. Then we may embed* ν *in* Y *if and only if* $m = \int_I s(y)\,\nu(dy)$ *exists, and* $m \geq 0$ *(resp.* $m \leq 0$*).*

(iii) $s(I)^\circ = (\alpha, \beta)$*,* $\alpha, \beta \in \mathbb{R}$*. Then we may embed* ν *in* Y *if and only if* $m = 0$*.*

The statement of the result in Pedersen and Peskir (2001) has the additional assumption in Case *(i)* that $\int_I |s(y)|\,\nu(dy) < \infty$. This can be dropped since in Case *(i)* the diffusion is recurrent so that either the 'Quick and Dirty' stopping time defined the introduction, or the extension of the Perkins embedding we introduced in the previous section, can be used to embed μ.

For the precise details of the proof of Lemma 2.10 we refer the reader to Pedersen and Peskir (2001). However we can provide a sketch of the proof using the modified Perkins embedding. For t less than the first exit time of the diffusion from the interior of $s(I)$ we have $M_t = s(Y_t) = B_{\tau_t}$ for some time-change τ and Brownian motion B. If $\langle M \rangle_\infty = \tau_\infty < \infty$ we may extend the time domain on which B_{τ_t} is defined to all positive times by continuing the Brownian motion beyond τ_∞. In this way we may drop the assumption of Theorem 2.4 that the process M_t has infinite variation. We deduce that we may embed our distribution on $s(I)^\circ$ if and only if, when we consider the problem of embedding μ in Brownian motion, our process remains on $s(I)^\circ$. However the transformed target distribution has support concentrated only on this interval, so when we consider the stopping time T defined in (2.22) and the form of $\gamma_+(\lambda)$ and $\gamma_-(\lambda)$ in the martingale scale, we see that problems can only occur if $\gamma_+(\lambda) = \infty$ or $\gamma_-(\lambda) = \infty$ for some λ. Further examination shows that this is only possible when μ is not integrable, or not centred — see Remark 2.2 — and the three cases of Lemma 2.10 all follow.

Our aim in the remainder of this section is to look at some of the properties of the construction, and of embeddings in general. Our principal question is (c.f. Perkins (1986) and Jacka (1988), where the law of $\sup |Y_t|$ in the Brownian case with centred target distribution is considered),

> given a diffusion Y_t, and a law ν, when does there exists an embedding for which the law of the maximum modulus of the process, $\sup_t |Y_{T\wedge t}|$, lies in the space $\mathcal{L}^p$ of random variables with finite p^{th} moment?

Before answering this question we show how the results of the previous section can be used to define an embedding of a target law in a diffusion.

Given ν and $(Y_t)_{t\geq 0}$ define μ and $M = s(Y)$ as above. As before, for M on $s(I)$ we can define

$$\kappa_M(x) = \begin{cases} \int_{\{u\geq 0\}} (x \wedge u)\, \mu(du) & : \quad x \geq 0; \\ \int_{\{u<0\}} (|x| \wedge |u|)\, \mu(du) & : \quad x < 0, \end{cases}$$

together with the quantities defined in (2.6)–(2.11). Write

$$\kappa_Y(y) = \kappa_M(s(y)) = \begin{cases} \int_{\{w\geq 0\}} (s(y) \wedge s(w))\, \nu(dw) & : \quad y \geq 0; \\ \int_{\{w<0\}} (|s(y)| \wedge |s(w)|)\, \nu(dw) & : \quad y < 0. \end{cases}$$

and, for $z > 0$, define the quantities:

$$\rho_+(z) = \operatorname*{argmin}_{y>0} \left\{ \frac{\kappa_Y(z) - \kappa_Y(-y)}{s(z) - s(-y)} \right\}, \tag{2.27}$$

$$\rho_-(z) = \operatorname*{argmax}_{y>0} \left\{ \frac{\kappa_Y(y) - \kappa_Y(-z)}{s(y) - s(-z)} \right\}, \tag{2.28}$$

$$\zeta_+(z) = -\inf_{y>0} \left\{ \frac{\kappa_Y(z) - \kappa_Y(-y)}{s(z) - s(-y)} \right\}, \tag{2.29}$$

$$\zeta_-(z) = \sup_{y>0} \left\{ \frac{\kappa_Y(y) - \kappa_Y(-z)}{s(y) - s(-z)} \right\}, \tag{2.30}$$

$$\nu_+(z) = \zeta_+(z) + \nu([z, \infty)), \tag{2.31}$$

$$\nu_-(z) = \nu((-\infty, -z]) + \zeta_-(z). \tag{2.32}$$

By convention, if $\rho_+(z)$ or $\rho_-(z)$ is not uniquely defined then we take the smallest solution.

Now define a stopping time for Y_t by:

$$\begin{aligned} T &= \inf\{t > 0 : Y_t \notin (-\rho_+(\overline{Y}_t), \rho_-(-\underline{Y}_t))\} \\ &= \inf\{t > 0 : M_t \notin (-\gamma_+(\overline{M}_t), \gamma_-(-\underline{M}_t))\}. \end{aligned} \tag{2.33}$$

The two alternative characterisations of T are equivalent because of the identities

$$\begin{aligned} s(-\rho_+(z)) &= -\gamma_+(s(z)), \\ s(\rho_-(z)) &= \gamma_-(-s(-z)). \end{aligned}$$

We also have that $\zeta_+(z) = \theta_+(s(z))$, and $\zeta_-(z) = \theta_-(-s(-z))$. It follows that T embeds

μ in $(M_t)_{t\geq 0}$, and hence ν in $(Y_t)_{t\geq 0}$. Also ν_+ and ν_- are the laws of the supremum and infimum respectively of $Y_{T\wedge t}$. Consequently we may restate Theorem 2.4 in the diffusion context.

Theorem 2.11. *Let $(Y_t)_{t\geq 0}$ be a regular, time-homogeneous diffusion, vanishing at zero and with supremum process $\overline{Y}_t$ and infimum process $\underline{Y}_t$, and let T be a stopping time such that $Y_T \sim \nu$. Then, for all $\lambda \geq 0$, the following hold:*

$$\mathbb{P}(\overline{Y}_T \geq \lambda) \geq \nu_+(\lambda), \tag{2.34}$$

$$\mathbb{P}(\underline{Y}_T \leq -\lambda) \geq \nu_-(\lambda). \tag{2.35}$$

If there exists an embedding, the stopping time T defined in (2.33) *is an embedding and is optimal in the sense that it attains equality in* (2.34) *and* (2.35)*.*

We are interested in the measure ν_* where ν_* is the law of $\sup_{t\leq T}|Y_t|$. Trivially, for $z \geq 0$,

$$\max\left(\nu_+(z), \nu_-(z)\right) \leq \nu_*([z,\infty)) \leq \nu_+(z) + \nu_-(z), \tag{2.36}$$

and it follows that $\nu_* \in \mathcal{L}^p$ if and only both ν_+ and ν_- are elements of $\mathcal{L}^p$.

The next two lemmas give upper and lower bounds on ν_+ and ν_-. We give proofs in the case of ν_+; the corresponding results for ν_- can be deduced using the transformation $\mu \mapsto \tilde{\mu}$.

Lemma 2.12. *For all $z > 0$, we have*

$$\begin{aligned}
\nu_+(z) \leq & \frac{1}{s(z)}\left[\kappa_Y(-z) - \kappa_Y(z) - |s(-z)|\nu((-\infty,-z])\right]_+ \mathbf{1}_{\{z>\rho_+(z)\}} \\
& + \nu(\{|y| \geq z\}),
\end{aligned} \tag{2.37}$$

$$\begin{aligned}
\nu_-(z) \leq & \frac{1}{|s(-z)|}\left[\kappa_Y(z) - \kappa_Y(-z) - s(z)\nu([z,\infty))\right]_+ \mathbf{1}_{\{z>\rho_-(z)\}} \\
& + \nu(\{|y| \geq z\}).
\end{aligned} \tag{2.38}$$

Proof. Suppose first that $z > \rho_+(z)$, or equivalently $s(-z) < -\gamma_+(s(z))$. Then by the concavity of κ_M on $\mathbb{R}_-$,

$$\kappa_M(-\gamma_+(s(z))) - \gamma_+(s(z))\theta_+(s(z)) \leq \kappa_M(s(-z)) + s(-z)\nu((-\infty,-z]),$$

which translates to

$$\kappa_Y(-\rho_+(z)) + s(-\rho_+(z))\zeta_+(z) \leq \kappa_Y(-z) + s(-z)\nu((-\infty,-z]).$$

Substituting this inequality into (2.29) we deduce that

$$\begin{aligned} s(z)\zeta_+(z) &= s(-\rho_+(z))\zeta_+(z) + \kappa_Y(-\rho_+(z)) - \kappa_Y(z) \\ &\leq \kappa_Y(-z) - \kappa_Y(z) + s(-z)\nu((-\infty, -z]). \end{aligned}$$

Conversely, if $z \leq \rho_+(z)$, then

$$\zeta_+(z) \leq \nu((-\infty, -\rho_+(z)]) \leq \nu((-\infty, -z]).$$

Given that $\nu_+(z) = \nu([z,\infty)) + \zeta_+(z)$, these two bounds lead directly to (2.37). $\square$

Lemma 2.13. *For all $z > 0$, we have*

$$\nu_+(z) \geq \frac{[\kappa_Y(-z) - \kappa_Y(z)]_+}{s(z) + |s(-z)|} + \nu([z,\infty)), \tag{2.39}$$

$$\nu_-(z) \geq \frac{[\kappa_Y(z) - \kappa_Y(-z)]_+}{s(z) + |s(-z)|} + \nu((-\infty, -z]). \tag{2.40}$$

Proof. By (2.29), for $z > 0$,

$$\zeta_+(z) \geq \frac{\kappa_Y(-z) - \kappa_Y(z)}{s(z) + |s(-z)|}.$$

Since also $\zeta_+(z) \geq 0$ the result follows easily from the identity $\nu_+(z) = \nu([z,\infty)) + \zeta_+(z)$. $\square$

Corollary 2.14. *For $z > 0$, we have:*

$$\begin{aligned} &\left(\frac{1}{s(z)} + \frac{1}{|s(-z)|}\right)|\kappa_Y(z) - \kappa_Y(-z)| + 2\nu(\{|y| \geq z\}) \\ &\qquad \geq \nu_+(z) + \nu_-(z) \geq \frac{|\kappa_Y(z) - \kappa_Y(-z)|}{s(z) + |s(-z)|} + \nu(\{|y| \geq z\}). \end{aligned}$$

Let T' be an embedding of ν in Y. For $p > 0$ we say this embedding is a H^p-embedding if $\sup_t |Y_{t\wedge T'}|$ is in $\mathcal{L}^p$. We may ask when does there exist a solution of the Skorokhod problem which is a H^p-embedding, and when is every ('sensible') solution of the Skorokhod problem a H^p-embedding? In this thesis we are interested in the first of these questions. By the extremality properties of our embedding T it is clear that there exists a H^p-embedding if and only if T is a H^p-embedding.

Corollary 2.14 can be used to give necessary and sufficient conditions for ν_* to be an element of $\mathcal{L}^p$. In particular, the following result follows easily from Corollary 2.14 and

(2.36).

Theorem 2.15. *Let Y_t be a regular diffusion and suppose that ν can be embedded in Y. Consider the embedding T of ν given in* (2.33)*. A sufficient condition for T to be a H^p-embedding is that $\nu \in \mathcal{L}^p$ and*

$$\int^{\infty} z^{p-1}\left(\frac{1}{s(z)} + \frac{1}{|s(-z)|}\right)|\kappa_Y(z) - \kappa_Y(-z)|\, dz < \infty. \tag{2.41}$$

Necessary conditions are that $\nu \in \mathcal{L}^p$ and

$$\int_0^{\infty} z^{p-1}\frac{|\kappa_Y(z) - \kappa_Y(-z)|}{s(z) + |s(-z)|}\, dz < \infty. \tag{2.42}$$

Remark 2.16. Note that in the symmetric case where $s(z) = -s(-z)$ then (2.41) and (2.42) are equivalent and Theorem 2.15 gives a necessary and sufficient condition for T to be a H^p-embedding.

We return to the problem of the existence of a H^p-embedding in the next section, and close this section with a further observation about the optimality of the embedding T.

Remark 2.17. Fix a measurable function $f : \mathbb{R} \to \mathbb{R}$ and let $(Y_t)_{t\geq 0}$ be a regular diffusion with $Y_0 = 0$ and ν a probability measure on $\mathbb{R}$. Then the embedding defined in (2.33) minimises the distribution of $\sup_{t\geq 0} f(Y_{t\wedge T'})$ over all stopping times T' such that $Y_{T'} \sim \nu$.

In particular the minimising choice of stopping time does not depend on the function f. This is in contrast with the problem of finding the Skorokhod embedding which maximises the law of $\sup_{t\geq 0} f(Y_{t\wedge T'})$. In that case the optimal embedding will in general depend on f.

2.4 H^p Embeddings for Diffusions.

Our goal in this section is to investigate further conditions on whether T is a H^p-embedding in the cases when $s(I)^\circ = (-\infty, \alpha), (\alpha, \infty), (\beta, \alpha)$ and $\mathbb{R}$. The first two cases are equivalent up to the map $x \mapsto -x$ and we consider them first.

2.4.1 Diffusions Transient to $+\infty$.

Theorem 2.18. *Let Y_t be a diffusion on I with scale function $s(z)$, such that $s(0) = 0$, $\sup_{z\in I} s(z) = \alpha < \infty$, and $\inf_{z\in I} s(z) = -\infty$. We may embed a law ν in Y if and only if $\int_I |s(z)|\,\nu(dz) < \infty$ and $m = \int_I s(z)\,\nu(dz) \geq 0$.*

Under these conditions:

- *if $m > 0$, then a necessary and sufficient condition for $\mathbb{E}\sup_t |Y_{T\wedge t}|^p < \infty$ is that*

$$\int^{\infty} \frac{z^{p-1}}{|s(-z)|}\,dz < \infty \text{ and } \nu \in \mathcal{L}^p; \tag{2.43}$$

- *if $m = 0$, this is also a sufficient condition. A necessary and sufficient condition is:*

$$\int^{\infty} \frac{z^{p-1}}{|s(-z)|}|\kappa_Y(z) - \kappa_Y(-z)|\,dz < \infty \text{ and } \nu \in \mathcal{L}^p. \tag{2.44}$$

Proof. The first part of this Theorem is a restatement of Lemma 2.10*(ii)* (or equivalently Pedersen and Peskir (2001)[Theorem 2.1]). For the second part assume $m \geq 0$ where $m = \int_0^\infty s(y)\,\nu(dy) - \int_{-\infty}^0 |s(y)|\,\nu(dy)$. For $z \geq 0$,

$$\begin{aligned}
\kappa_Y(-z) - \kappa_Y(z) = &- \int_{\{y<-z\}} |s(y)|\,\nu(dy) + \int_{\{y>z\}} s(y)\,\nu(dy) - m \\
&+ \int_{\{y\leq -z\}} |s(-z)|\,\nu(dy) - \int_{\{y\geq z\}} s(z)\,\nu(dy) \\
\leq &\int_{\{y>z\}} s(y)\,\nu(dy) + \int_{\{y\leq -z\}} |s(-z)|\,\nu(dy),
\end{aligned}$$

so by Lemma 2.12,

$$\begin{aligned}
\nu_+(z) \leq &\frac{1}{s(z)} \left[\kappa_Y(-z) - \kappa_Y(z) - |s(-z)|\nu((-\infty, -z])\right]_+ \mathbf{1}_{\{z>\rho_+(z)\}} \\
&+ \nu(\{|y| \geq z\}) \\
\leq &\int_{\{y>z\}} \frac{s(y)}{s(z)}\,\nu(dy) + \nu(\{|y| \geq z\}) \\
\leq &\frac{\alpha}{s(z)}\nu(\{|y| \geq z\}).
\end{aligned}$$

Since $\alpha/s(z) < 2$ for sufficiently large z it follows that $\nu \in \mathcal{L}^p$ is a necessary and sufficient condition for $\nu_+ \in \mathcal{L}^p$.

Now consider $\nu_-(z)$. We note that given $\varepsilon > 0$, for sufficiently large z,

$$m - \varepsilon \le \kappa_Y(z) - \kappa_Y(-z) \le m + \varepsilon,$$

and so by Lemma 2.12,

$$\nu_-(z) \le \frac{1}{|s(-z)|}(m+\varepsilon) + \nu(\{|y| \ge z\}).$$

As a result (2.43) is a sufficient condition for $\nu_- \in \mathcal{L}^p$ when $m \ge 0$.

Conversely, if $m > 0$ Lemma 2.13 implies that for sufficiently large z,

$$\nu_-(z) \ge \frac{1}{2|s(-z)|}(m - \varepsilon),$$

and so (2.43) is also necessary.

Now suppose $m = 0$. By (2.38),

$$\nu_-(z) \le \frac{1}{|s(-z)|}[\kappa_Y(z) - \kappa_Y(-z)]_+ + \nu(\{|y| \ge z\})$$

so (2.44) is a sufficient condition for $\nu_- \in \mathcal{L}^p$. By Corollary 2.14, for sufficiently large z,

$$\nu_+(z) + \nu_-(z) \ge \frac{|\kappa_Y(z) - \kappa_Y(-z)|}{2|s(-z)|} + \nu(\{|y| \ge z\}).$$

If $\nu_* \in \mathcal{L}^p$ then both ν_+ and ν_- lie in $\mathcal{L}^p$, and so (2.44) is a necessary condition.

□

Example 2.19 (Drifting Brownian Motion)**.** Suppose Y is drifting Brownian motion on $\mathbb{R}$,

$$Y_t = B_t + \varphi t,$$

for $t \ge 0$ and $\varphi > 0$. Then $s(y) = 1 - \mathrm{e}^{-2\varphi y}$ is the scale function for Y, so $\sup_y s(y) = 1$. If $\int_{\mathbb{R}} s(y)\,\nu(dy) < 0$, then it is not possible to embed ν in Y. If $\int_{\mathbb{R}} s(y)\,\nu(dy) \ge 0$, we may embed ν in Y, and since

$$\int^{\infty} \frac{y^{p-1}}{|s(-y)|}\,dy = \int^{\infty} \frac{y^{p-1}}{\mathrm{e}^{2\varphi y} - 1}\,dy < \infty,$$

if follows that if $\nu \in \mathcal{L}^p$, then $\sup_t |Y_{T\wedge t}|$ is too.

These conclusions should be compared with those in Grandits and Falkner (2000).

Grandits and Falkner conclude that if Y is drifting Brownian motion, and if T' is any embedding of an integrable target distribution ν in Y, then $T' \in H^1$.

Example 2.20 (Bessel d Process)**.** In Hambly et al. (2003) the authors consider a Skorokhod embedding for the $\text{B}\textsc{es}(3)$ process. For $d > 2$ let Y solve

$$dY_t = dB_t + \frac{d-1}{2Y_t}dt, \qquad\qquad Y_0 = 1.$$

Then $I = (0, \infty)$ and $s(y) = -y^{2-d}$. We do not have $Y_0 = 0$, nor $s(0) = 0$ but the modifications to the theory are trivial. We can embed ν in Y if and only if $\int_0^\infty y^{2-d}\nu(dy) < 1$. Furthermore Y is only defined on the positive reals, so in deciding whether $\nu_* \in \mathcal{L}^p$ we need only consider ν_+. But, provided we may embed ν in Y, it follows from the proof of Theorem 2.18 that a necessary and sufficient condition for $\nu_+ \in \mathcal{L}^p$ is $\nu \in \mathcal{L}^p$.

2.4.2 Recurrent Diffusions

The general case is covered by Theorem 2.15. If we have some control on the scale function then we are able to make the results more explicit.

Theorem 2.21. *Suppose for $|y| \geq 1$ there exists $k, K > 0$ such that*

$$k|y|^r \leq |s(y)| \leq K|y|^q, \textit{ for some } q \geq r \geq 0. \tag{2.45}$$

Then for $p > 0$,

(i) if $p > q$,

$$m = 0 \textit{ and } \nu \in \mathcal{L}^{p+q-r} \implies \nu_* \in \mathcal{L}^p \implies \nu \in \mathcal{L}^p \textit{ and } m = 0;$$

(ii) if $p < r$,

$$\nu \in \mathcal{L}^{p+q-r} \implies \nu_* \in \mathcal{L}^p \implies \nu \in \mathcal{L}^p;$$

(iii) if $r \leq p \leq q$,

$$\int_1^\infty y^{p-r-1}|\kappa_Y(y) - \kappa_Y(-y)|\,dy < \infty \textit{ and } \nu \in \mathcal{L}^p \tag{2.46}$$

$$\implies \nu_* \in \mathcal{L}^p$$

$$\implies \nu \in \mathcal{L}^p \textit{ and } \int_0^\infty y^{p-q-1}|\kappa_Y(y) - \kappa_Y(-y)|\,dy < \infty. \tag{2.47}$$

In particular, if $r = q$, the three cases each become if and only if statements.

Remark 2.22. The case where the diffusion is in natural scale, so that $s(y) = y$, is the case considered by Perkins (1986). Here the Cases *(i)* and *(ii)* are dealt with in his introduction, while in Case *(iii)* he shows that $\nu \in \mathcal{L}^1$, $m = 0$ and $H(\mu) < \infty$, where

$$H(\mu) = \int_0^\infty y^{-1} \left| \int_{-\infty}^\infty x \mathbf{1}_{\{|x| \geq y\}} \, \mu(dx) \right| \, dy,$$

are necessary and sufficient conditions for $\nu_* \in \mathcal{L}^1$. It is not hard to see that this condition is equivalent to (2.47).

Proof. *(i)* Suppose $p > q$. If $\nu \in \mathcal{L}^q$ then since $|s(y)| \leq K|y|^q$ for $|y| \geq 1$, we have $\int |s(y)| \, \nu(dy) < \infty$, so m exists.

Now suppose $m = 0$ and $\nu \in \mathcal{L}^{p+q-r}$. By Theorem 2.15 it is sufficient to show

$$\int_1^\infty y^{p-1} \left(\frac{1}{s(y)} + \frac{1}{|s(-y)|} \right) |\kappa_Y(y) - \kappa_Y(-y)| \, dy < \infty.$$

For $y > 0$,

$$\begin{aligned}
&\kappa_Y(y) - \kappa_Y(-y) \\
&\qquad = \int_{\{|w| \leq y\}} s(w) \, \nu(dw) + \int_{\{w > y\}} s(y) \, \nu(dw) - \int_{\{w < -y\}} |s(-y)| \, \nu(dw) \\
&\qquad = -\int_{\{|w| > y\}} s(w) \, \nu(dw) + s(y) \nu(\{w > y\}) - |s(-y)| \nu(\{w < -y\}),
\end{aligned}$$

where we have used the fact that $m = 0$. By assumption

$$\left(\frac{1}{s(y)} + \frac{1}{|s(-y)|} \right) \leq \frac{2}{ky^r}, \text{ for } y \geq 1$$

so that

$$\begin{aligned}
\int_1^\infty & y^{p-1} \left(\frac{1}{s(y)} + \frac{1}{|s(-y)|} \right) |\kappa_Y(y) - \kappa_Y(-y)| \, dy \\
&\leq \frac{2}{k} \int_1^\infty y^{p-r-1} \Big[K y^q \nu((y, \infty)) + K y^q \nu((-\infty, -y)) \\
&\qquad\qquad + \int_{\{|w| > y\}} |s(w)| \, \nu(dw) \Big] \, dy.
\end{aligned}$$

The first two terms in the bracket will be finite upon integration since $\nu \in \mathcal{L}^{p+q-r}$.

Also, by Fubini,

$$\int_1^\infty y^{p-r-1}\left[\int_{\{w>y\}} s(w)\,\nu(dw)\right]\,dy = \int_{\{w>1\}}\left[\int_1^w y^{p-r-1}s(w)\,dy\right]\,\nu(dw)$$
$$\leq K\int_{\{w>1\}}\frac{w^{q+p-r}}{p+q-r}\,\nu(dw)<\infty.$$

We can show a similar result for the integral over $\{w<-1\}$ and it follows that $\nu_*\in\mathcal{L}^p$.

Now suppose that $\nu_*\in\mathcal{L}^p$. Then clearly $\nu\in\mathcal{L}^p$, and

$$\mathbb{E}\sup_t|s(Y_{T\wedge t})|\leq K\mathbb{E}\left(\sup_t|Y_{T\wedge t}|^q+1\right)\leq K\mathbb{E}\left(\sup_t|Y_{T\wedge t}|^p\right)+K<\infty.$$

Furthermore $s(Y_t)$ is a local martingale, so, since $\mathbb{E}\sup_t|s(Y_{T\wedge t})|<\infty$, $s(Y_{T\wedge t})$ is a UI martingale, and hence

$$m=\mathbb{E}\left(s(Y_T)\right)=0.$$

(ii) Suppose now $p<r$, and $\nu\in\mathcal{L}^{p+q-r}$. Then as before, by Theorem 2.15 it is sufficient to show

$$\int_1^\infty y^{p-1}\left(\frac{1}{s(y)}+\frac{1}{|s(-y)|}\right)|\kappa_Y(y)-\kappa_Y(-y)|\,dy<\infty.$$

A simple inequality gives

$$\begin{aligned}|\kappa_Y(y)-\kappa_Y(-y)|&\leq\kappa_Y(y)+\kappa_Y(-y)\\&=\int_{\{|w|\leq y\}}|s(w)|\,\nu(dw)+s(y)\nu(\{w>y\})+|s(-y)|\nu(\{w<-y\}),\end{aligned}$$

and so

$$\begin{aligned}&\int_1^\infty\left(\frac{1}{s(y)}+\frac{1}{|s(-y)|}\right)|\kappa_Y(y)-\kappa_Y(-y)|\,dy\\&\quad\leq\frac{2}{k}\int_1^\infty y^{p-r-1}\Big[Ky^q\nu(\{|w|>y\})+\int_{\{|w|\leq y\}}|s(w)|\,\nu(dy)\Big]\,dy,\end{aligned}$$

where, as before, the first term is finite upon integration. For the final term

$$\begin{aligned}
\int_1^\infty y^{p-r-1}\left[\int_{\{0<w\le y\}} s(w)\,\nu(dw)\right]dy
&= \int_{\{w>0\}} s(w)\left[\int_{w\vee 1}^\infty y^{p-r-1}\,dy\right]\nu(dw)\\
&\le \int_{\{w>0\}} \frac{(w\vee 1)^{p-r}}{r-p} s(w)\,\nu(dw)\\
&\le \int_0^1 \frac{s(w)}{r-p}\,\nu(dw) + \frac{K}{r-p}\int_{\{w>1\}} w^{p+q-r}\,\nu(dw),
\end{aligned}$$

which is finite by assumption since $\nu \in \mathcal{L}^{p+q-r}$. The corresponding result also holds over $\{w<0\}$. So we have shown $\nu \in \mathcal{L}^{p+q-r} \implies \nu_* \in \mathcal{L}^p$. The second implication $\nu_* \in \mathcal{L}^p \implies \nu \in \mathcal{L}^p$ is clear.

(iii) This case is a trivial application of (2.45) to Theorem 2.15. ☐

For the integral condition in (2.46) to hold, a necessary condition is that $|\kappa_Y(z) - \kappa_Y(-z)| \to 0$ as $z \to \infty$. However this occurs if and only if $m = 0$, provided m exists. So if m exists, if $r = p = q$ and if $\nu \in \mathcal{L}^p$, then $m = 0$ is a necessary condition for $\nu_* \in \mathcal{L}^p$. We show in Example 2.23 that this condition is not sufficient.

Note that it is not necessary for m to exist for the integral condition in (2.42) to be satisfied, and for ν_* to be an element of $\mathcal{L}^p$. For example, suppose that both the scale function and the target measure are symmetric about 0, i.e. suppose $s(z) = -s(-z)$ and $\nu(dz) = \nu(d(-z))$. Then $\kappa_Y(z) = \kappa_Y(-z)$ and (2.42) is trivially satisfied. If s and ν are symmetric then $\nu_* \in \mathcal{L}^p$ if and only if $\nu \in \mathcal{L}^p$.

Example 2.23. We now consider a diffusion on $\mathbb{R}$ with behaviour specified by

$$dY_t = 2\sqrt{|Y|_t}dB_t + \delta\,\mathrm{sign}(Y_t)dt,$$

where $Y_0 = 0$, and $\delta \in (0,2)$. The solution to this SDE is not unique in law, but we make it so by assuming the law of the process is symmetric about 0, and that the process does not wait at 0. In particular, $|Y_t|$ is a Bessel process of dimension δ. Such a process is recurrent, and we can construct the process Y_t from $|Y_t|$ by assigning to each excursion away from 0 an independent random variable with value either 1 or -1. Alternatively we may define the process by its scale function

$$s(y) = (|y|^{1-\frac{\delta}{2}})\,\mathrm{sign}(y),$$

and write $Y_t = s(W_{\tau_t})$, for a Brownian motion W_t and a suitable time change τ_t. Since $(Y_t)_{t\geq 0}$ is recurrent on $\mathbb{R}$ we may embed any target distribution.

We may apply Theorem 2.21 to this process for some target distribution ν and examine the behaviour of $\sup_t |Y_{T\wedge t}|$, for our embedding T. We note that, using the notation of Theorem 2.21, $r = q = 1 - \frac{\delta}{2}$, so the statements in the theorem each become if and only if statements. We can consider each case separately:

(i) In the case where $p > 1 - \frac{\delta}{2}$, $\nu \in \mathcal{L}^p$ guarantees that m exists, and a necessary and sufficient condition for $\sup_t |Y_{T\wedge t}| \in \mathcal{L}^p$ is that $m = 0$.

(ii) If $p < 1 - \frac{\delta}{2}$, $\nu \in \mathcal{L}^p$ is both necessary and sufficient for $\sup_t |Y_{T\wedge t}| \in \mathcal{L}^p$.

(iii) Suppose now that $p = 1 - \frac{\delta}{2}$. If $m \neq 0$ then $\sup_t |Y_{T\wedge t}| \notin \mathcal{L}^p$. However we now show that $m = 0$ is not a sufficient condition for $\sup_t |Y_{T\wedge t}| \in \mathcal{L}^p$.

We embed the probability measure ν defined by

$$\nu(dy) = \frac{y^{-p-1}}{(\log y)^2}\, dy \qquad\qquad \text{for } y \geq \mathrm{e},$$

with the rest of the mass placed at $-b$. Here b is chosen such that $\int s(y)\,\nu(dy) = 0$. It can be checked that $\nu \in \mathcal{L}^p$. Then, provided $z > \max(\mathrm{e}, -s^{-1}(-b))$,

$$\begin{aligned}|\kappa_Y(z) - \kappa_Y(-z)| &= \int_z^\infty \frac{1}{y(\log y)^2}\, dy - z^p \nu((z,\infty)) \\ &= \frac{1}{\log z} - z^p \nu([z,\infty)).\end{aligned}$$

Consequently, because $\nu \in \mathcal{L}^p$ and $\int_z^\infty \frac{1}{y\log(y)}\, dy = \infty$,

$$\int^\infty y^{-1}|\kappa_Y(y) - \kappa_Y(-y)|\, dy = \infty.$$

So $m = 0$ is not sufficient to ensure that $\sup_t |Y_{T\wedge t}| \in \mathcal{L}^p$.

2.4.3 Diffusions Which in Natural Scale Have State Space Consisting of a Finite Interval.

Theorem 2.24. *Let Y_t be a diffusion on I with scale function $s(z)$, such that $s(0) = 0$, $\sup_{z\in I} s(z) = \alpha < \infty$, and $\inf_{z\in I} s(z) = \beta > -\infty$. We may embed a law ν in Y if and only if $\int_I |s(z)|\,\nu(dz) < \infty$ and $m = \int_I s(z)\,\nu(dz) = 0$.*

Furthermore $\nu_* \in \mathcal{L}^p$ *if and only if* $\nu \in \mathcal{L}^p$.

Proof. The first part of this result follows from Lemma 2.10*(iii)* (or equivalently Pedersen and Peskir (2001)[Theorem 2.1]). The remaining part follows from Theorem 2.21. In our setting the scale function s is bounded — so we have $q = r = 0$, $p > 0$ and we are in case *(i)* of Lemma 2.10. In particular, m exists, and $\nu_* \in \mathcal{L}^p$ if and only if $m = 0$ and $\nu \in \mathcal{L}^p$. However we have already noted that in order to be able to embed in this case we must have $m = 0$, so our condition is essentially $\nu_* \in \mathcal{L}^p \iff \nu \in \mathcal{L}^p$. $\square$

Chapter 3

Minimality and the Azema-Yor Solution

We now turn to considering embeddings which *maximise* the distribution of the maximum. In the centred case the solution to this problem, among the class of uniformly integrable embeddings, is known to be the embedding described in Azéma and Yor (1979*a*). We begin the chapter with a review of this embedding.

When we consider non-centred target distributions it is no longer appropriate to consider the class of uniformly integrable martingales. Instead we propose using the class of minimal stopping times of Monroe (1972). In particular we deduce necessary and sufficient conditions for a stoppping time embedding a non-centred distribution to be minimal in terms of properties of the local-martingale $B_{t\wedge T}$. Using this equivalence we are able to extend the Azema-Yor embedding to non-centred target laws and show that the extension retains the optimality property of the original embedding.

Finally we show that these ideas extend naturally to diffusions, and that we are able to extend an idea of Jacka (1988) to find embeddings which maximise $\sup_{t\leq T}|Y_t|$ and more generally $\sup_{t\leq T} f(Y_t)$ for a general function f.

3.1 Introduction

The work we present in this chapter is motivated by the following question:

> Given a diffusion $(X_t)_{t\geq 0}$ and a target distribution μ_X for which an embedding exists, which embedding maximises the law of $\sup_{s\leq T} X_s$ (respectively $\sup_{s\leq T} |X_s|$)?

For Brownian motion, the question has been solved by Azéma and Yor (1979*a*) (respectively Jacka (1988)) in the class of stopping times for which $B_{t\wedge T}$ is a UI-martingale.

There are several considerations that need to be made when moving from the Brownian case to the diffusion case. Firstly, the mean-zero assumption that is made by Azéma and Yor (1979*a*) and Jacka (1988) is no longer natural since we are no longer necessarily dealing with a martingale. The second aspect that needs to be considered is with what restriction should we replace the UI condition? That such a condition is desirable may be seen by considering a recurrent diffusion. Here the maximisation problem can easily seen to be degenerate by considering first running the diffusion until it hits a level x, allowing it to return to the origin and then using the reader's favourite embedding. Clearly this dominates the unmodified version of the reader's favourite embedding.

In Pedersen and Peskir (2001) an integrability condition on the maximum (specifically that $\mathbb{E}(\sup_{s\leq T} s(X_s)) < \infty$ where s is the scale function of X) was suggested to replace the UI condition in the Brownian case. In this work we propose using the following class of stopping times introduced by Monroe (1972) to provide us with a natural restriction on the set of admissible embeddings.

Definition 3.1. A stopping time T for the process X is *minimal* if whenever $S \leq T$ is a stopping time such that X_S and X_T have the same distribution then $S = T$ a.s..

The class of minimal stopping times provides us with a natural link to the uniformly integrable Brownian case as a consequence of the following result:

Theorem 3.2. *(Monroe, 1972, Theorem 3) Let S be a stopping time such that $\mathbb{E}(B_S) = 0$. Then S is minimal if and only if the process $B_{t\wedge S}$ is uniformly integrable.*

It will turn out that the minimality idea fits well with the problem of embedding in diffusions. As in the previous chapter, our approach to embedding in diffusions will be to map the diffusion into natural scale (so that, up to a time change, it resembles Brownian motion) and use techniques developed for embedding Brownian motion.

Using this method on a transient diffusion one finds that the state space and target distribution for the Brownian motion is restricted to a half-line (or sometimes a finite interval). We will show minimality to be equivalent to stopping the Brownian motion before it leaves this interval, so that a minimal stopping time is necessarily before the explosion time of X.

When we map from the problem of embedding μ_X in X to the Brownian scale the target law μ we obtain for B is the image of μ_X under the scale function. The key point is that there is no reason why this target law should have mean zero. Thus, unlike most of the other studies of Skorokhod embeddings in Brownian motion we are interested in non-centred target distributions, and non-UI stopping times. One of our main results is to recharacterise the minimality condition on T in terms of a condition on $\mathbb{E}(B_T|\mathcal{F}_S)$ for stopping times $S \leq T$. In fact most of the chapter will concentrate on embedding non-centred target distributions in B, and we will only return to the diffusion case in a short final section.

The chapter will proceed as follows. In Section 3.2 we construct the classical Azema-Yor embedding (see Azéma and Yor (1979a)) to introduce the reader to the construction we will use later. Then in Section 3.3 we prove some results concerning minimality of stopping times for non-centred target distributions, giving an equivalent condition to minimality in terms of the process. In particular, given a non-minimal embedding T, we show in Section 3.4 how to construct a new (minimal) stopping time $T' \leq T$ which embeds μ. Next, in Section 3.5 we construct an extension of the Azema-Yor embedding for non-centred target distributions and show both that it is minimal, and that it retains the optimality properties of the original Azema-Yor embedding. In Section 3.6 we use these stopping times to construct an embedding maximising the distribution of $\sup_{s \leq T} h(B_s)$ for a general function h. Finally in Section 3.7 we apply these results to the problem of embedding optimally in diffusions.

Throughout this chapter we work with a standard Brownian motion with $B_0 = 0$. In the next chapter we will want to consider the case where our basic process is a Brownian motion with general starting law, $B_0 \sim \mu_0$ and in this case it will turn out that some of work in this chapter will be relevant to the later problem. However for clarity of exposition we ignore the general case for this chapter and we will consider when the results can be extended in Chapter 4.

3.2 The Azema-Yor Embedding

We begin by introducing the Azema-Yor embedding (see Azéma and Yor (1979a)) and a notation which we will use in later sections.

Let μ be our target distribution on $\mathbb{R}$, with mean $m = 0$, and let B_t be a Brownian motion with $B_0 = 0$. Our goal is to embed μ in B.

Define[1]

$$\eta(x) := \mathbb{E}^{\mu}|X - x|. \tag{3.1}$$

The definition of η ensures that $\eta(x)$ is a convex function which is asymptotic to, and greater than or equal to, the function $|x|$. For $\theta \in [-1, 1]$,

$$u(\theta) := \inf\{y \in \mathbb{R} : \eta(y) + \theta(x - y) \leq \eta(x), \forall x \in \mathbb{R}\}. \tag{3.2}$$

We will later want to use the inverse function, which we will define to be

$$u^{-1}(y) = \inf\{\theta : u(\theta) \geq y\}.$$

For an interpretation of these and subsequent quantities we refer the reader to Figure 3-1. Our interpretation of θ in (3.2) is that it is the gradient of a tangent to η, and then $u(\theta)$ is the smallest x at which there exists a tangent to η with gradient θ.

Let

$$z_+(\theta) := \frac{\eta(u(\theta)) - \theta u(\theta)}{1 - \theta}, \tag{3.3}$$

and define also

$$b(w) := u(z_+^{-1}(w))$$

for $0 \leq w \leq \sup\{\text{supp}(\mu)\}$. The function b is well defined and left-continuous since $z_+(\theta)$ is a continuous bijection $z_+ : [-1, 1] \to [0, \sup\{\text{supp}(\mu)\}]$ (if $\sup\{\text{supp}(\mu)\} = \infty$, $u(1) = \infty$ and we take $z_+(1) = \infty$). We interpret $z_+(\theta)$ as the x-co-ordinate of the intersection of the line $y = x$ and the tangent with gradient θ. It follows that $b(w)$ is the x-value of the left-most point on $(x, \eta(x))$ with the property that the tangent through this point hits the line $y = x$ at w, see Figure 3-1 for a pictorial representation of this idea.

Lemma 3.3 (The Azema-Yor Embedding)**.** *For μ, B as above, define the stopping*

[1] This function is related to the function κ introduced in Chapter 2, and also to the potential of the measure μ. The exact relationship is given in (4.3).

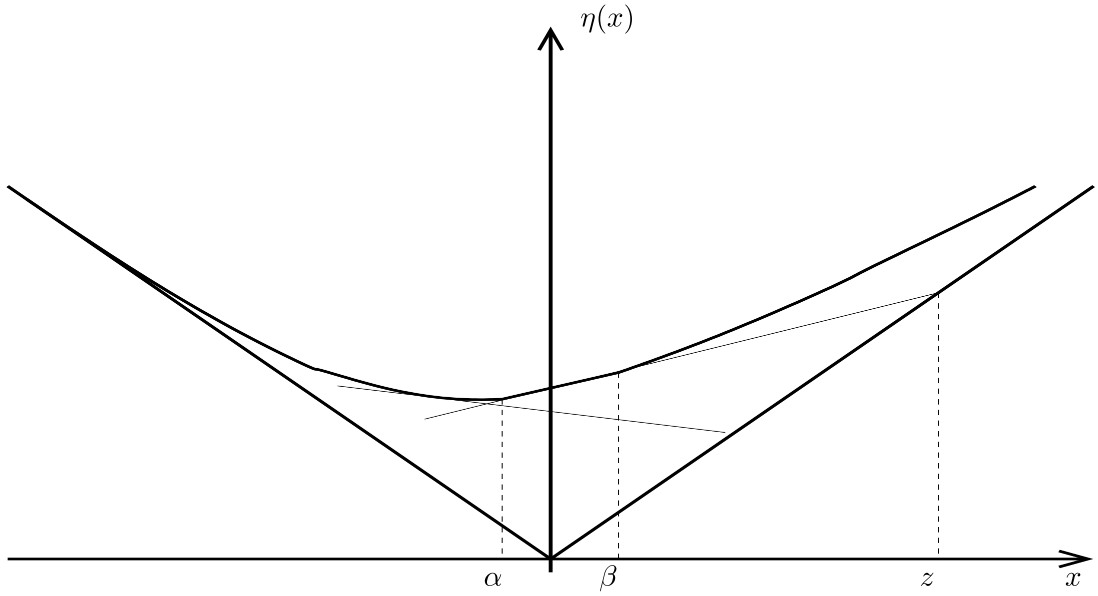

Figure 3-1: $\eta(x)$ where the support of the corresponding distribution μ is bounded below. Further $\mu((\alpha,\beta)) = 0$ and there is an atom at β. Consequently there are multiple tangents to η at β. We have that $\alpha = b(z)$. Note that $u(u^{-1}(\beta)) = \alpha$, but that $\eta(\alpha) - \alpha u^{-1}(\alpha) = \eta(\beta) - \beta u^{-1}(\beta)$.

time

$$T_{AY} := \inf\{t > 0 : B_t \leq b(\overline{B}_t)\}. \tag{3.4}$$

Then T_{AY} is a minimal embedding of μ in B. Further T_{AY} has the property that if T is another minimal stopping time which embeds μ in B, then

$$\mathbb{P}(\overline{B}_T \geq y) \leq \mathbb{P}(\overline{B}_{T_{AY}} \geq y). \tag{3.5}$$

Remark 3.4. The above theorem is the standard statement of the Azema-Yor embedding (e.g. Perkins (1986, Theorem 2.5)) except for the minimality condition on T. In other statements of the result T is required to be a stopping time for which the process $B_{T\wedge t}$ is a UI-martingale. However we are able to replace this condition with a minimality condition due to Theorem 3.2.

Classically (see e.g. Pedersen and Peskir (2001)), the above result is stated for the stopping time

$$T'_{AY} := \inf\{t > 0 : \overline{B}_t \geq \Psi(B_t)\},$$

where Ψ is the *barycentre* function:

$$\Psi(x) = \begin{cases} \frac{1}{\mu([x,\infty))} \int_{\{u \geq x\}} u\, \mu(du) & \mu([x,\infty)) > 0 \\ x & \mu([x,\infty)) = 0. \end{cases}$$

Here, we do not prove the result but show simply that $T_{AY} = T'_{AY}$.

Since

$$\begin{aligned} T_{AY} &= \inf\{t > 0 : B_t \leq b(\overline{B}_t)\} \\ &= \inf\{t > 0 : b^{-1}(B_t) \leq \overline{B}_t\} \end{aligned}$$

where $b^{-1}(y) = \inf\{z : b(z) \geq y\}$, it is sufficient to show that $\Psi = b^{-1}$. Note that $b^{-1}(y) = z_+(u^{-1}(y))$. Further, from the definition of η, we have that

$$\begin{aligned} u^{-1}(y) &= \eta'_-(y) \\ &= 1 - 2\mu([y,\infty)), \end{aligned} \tag{3.6}$$

where η'_- denotes the left-derivative of η (which exists by the convexity of η).

Now $u(u^{-1}(y)) = y$ unless there exits some $z < y$ for which $\mu((z,y)) = 0$, in which case however it is still true that $\eta(u(u^{-1}(y))) - u(u^{-1}(y))u^{-1}(y) = \eta(y) - yu^{-1}(y)$, since $\eta'(w) = \eta'_-(y)$ for all $w \in (z,y)$. Collecting all these observations together we have

$$\begin{aligned} b^{-1}(y) &= z_+(u^{-1}(y)) \\ &= \frac{\eta(u(u^{-1}(y))) - u(u^{-1}(y))u^{-1}(y)}{1 - u^{-1}(y)} \\ &= \frac{\eta(y) - yu^{-1}(y)}{1 - u^{-1}(y)} \\ &= \frac{\int |w - y|\, \mu(dw) - y(1 - 2\mu([y,\infty)))}{2\mu([y,\infty))} \\ &= \frac{1}{\mu([y,\infty))} \int_{\{w \geq y\}} w\, \mu(dw) \\ &= \Psi(y), \end{aligned} \tag{3.7}$$

and we are done.

Remark 3.5. Thus we have shown that the embedding we define in (3.4) is a representation of the Azema-Yor embedding and hence, by the well known properties of this embedding established in Blackwell and Dubins (1963), is optimal in the sense that it maximises the distribution of the maximum among all stopping times embedding μ in

B which are minimal (see also Azéma and Yor (1979*b*)).

3.3 Minimal Embeddings for Non-centred Distributions

In this section we examine the properties of minimal stopping times. In particular, we aim to find equivalent conditions to minimality — in a similar way to Theorem 3.2 — when the target distribution is not centred.

We begin by noting the following result from Monroe (1972) which justifies the existence of minimal stopping times:

Proposition 3.6 (Monroe (1972), Proposition 2)**.** *For any stopping time T there exists a minimal stopping time $S \leq T$ such that $B_S \sim B_T$.*

For completeness, we repeat the proof given in Monroe (1972).

Proof. Consider the class $\mathcal{T}$ of stopping times S which embed μ and $S \leq T$. There is a natural ordering on this set ($S_1 \preceq S_2$ iff $S_1 \leq S_2$ a.s.). Set $\alpha = \sup_{S \in \mathcal{T}} \mathbb{E}\mathrm{e}^{-T} \in (0,1]$. We can therefore find a sequence $S_1, S_2, \ldots$ of stopping times, decreasing in the natural ordering, such that $\mathbb{E}\mathrm{e}^{-S_n} \uparrow \alpha$. Then $S_n \downarrow S$ and the stopping time S is minimal, embeds μ and $S \leq T$. $\square$

Of course the above proof does not help us to construct a minimal stopping time, and the sequence chosen is not unique — there can be multiple minimal stopping times which are smaller than a given embedding. We shall see in Section 3.4 that we are able to provide an (essentially) constructive method for providing such stopping times.

The main result of this section is the following:

Theorem 3.7. *Let T be a stopping time of Brownian motion which embeds a distribution μ where $m = \int_{\mathbb{R}} x \, \mu(dx) < 0$. Then the following conditions are equivalent:*

(i) T is minimal for μ;

(ii) for all stopping times $R \leq S \leq T$,

$$\mathbb{E}(B_S | \mathcal{F}_R) \leq B_R \quad \text{a.s.;}$$

(iii) for all stopping times $S \leq T$,

$$\mathbb{E}(B_T | \mathcal{F}_S) \leq B_S \quad \text{a.s.;} \tag{3.8}$$

(iv) for all $\gamma > 0$

$$\mathbb{E}(B_T; T > H_{-\gamma}) \leq -\gamma \mathbb{P}(T > H_{-\gamma}),$$

where $H_\alpha = \inf\{t > 0 : B_t = \alpha\}$ is the hitting time of the level α;

(v) as $\gamma \to \infty$

$$\gamma \mathbb{P}(T > H_{-\gamma}) \to 0.$$

In the case where supp(μ) $\subseteq [\alpha, \infty)$ for some $\alpha < 0$ then the above conditions are also equivalent to the condition:

(vi)

$$\mathbb{P}(T \leq H_\alpha) = 1. \tag{3.9}$$

Remark 3.8. Of course the Theorem may be restated in the case where $m > 0$ by considering the process $-B_t$. We will use this observation extensively in Section 3.5.

Remark 3.9. Equation (3.8) makes us suspect that when T is minimal, the process $B_{t \wedge T}$ is in fact a supermartingale. To check this we need to show also that $\mathbb{E}B_{t\wedge T}^- < \infty$ for all t, where, for a random variable X we define $X^+ = X \vee 0$ and $X^- = (-X) \vee 0$ — the positive and negative parts respectively. We show this more generally, for a stopping time $S \leq T$. Using *(ii)*,

$$\mathbb{E}(B_T; B_T \leq 0) \leq \mathbb{E}(B_T; B_S \leq 0) \leq \mathbb{E}(B_S; B_S \leq 0),$$

so that $\mathbb{E}B_S^- < \mathbb{E}B_T^- < \infty$ and the process is indeed a supermartingale.

As a consequence of *(ii)*, if $S \leq T$ is a stopping time and T is minimal, then S is minimal too provided $\mathbb{E}B_S < 0$ and $\mathbb{E}|B_S| < \infty$. The first condition is a trivial consequence of *(ii)* on taking $R = 0$, the second condition then follows from the first on noting that $\mathbb{E}B_S^- < \infty$, $\mathbb{E}B_S = \mathbb{E}B_S^+ - \mathbb{E}B_S^-$ and $\mathbb{E}|B_S| = \mathbb{E}B_S^+ + \mathbb{E}B_S^-$.

Consequently we have the following corollary of Theorem 3.7:

Corollary 3.10. *If T is minimal and $S \leq T$ for a stopping time S then S is minimal for $\mathcal{L}(B_S)$.*

Remark 3.11. The third condition of Theorem 3.7 can be thought of as analogous to the condition

$$\forall S \leq T, \qquad \mathbb{E}(B_T|\mathcal{F}_S) = B_S \quad \text{a.s.}$$

in the case where $m = 0$. The proof of the corresponding result in Monroe (1972) shows that this is the key idea in showing that uniform integrability is equivalent to minimality. In the case we are interested in there is no equivalent notion to correspond with uniform integrability, so we use (3.8) instead.

Before the proof of Theorem 3.7 we prove the following, which (although very similar to the conclusions of Remark 3.9) will be necessary to show that condition *(v)* implies *(ii)*.

Proposition 3.12. *If T is a stopping time such that $B_T \sim \mu$ where μ is integrable and $m < 0$ and*

$$\gamma\mathbb{P}(T > H_{-\gamma}) \to 0, \tag{3.10}$$

as $\gamma \to \infty$ then $\mathbb{E}|B_S| < \infty$ and $\mathbb{E}B_S \le 0$ for all stopping times $S \le T$.

Proof of Proposition 3.12. We show that, for $S \le T$, $\mathbb{E}B_S^- < \infty$ and $\mathbb{E}B_S^+ \le \mathbb{E}B_S^-$ from which the result follows.

Suppose $\gamma > 0$. Since $B_{t\wedge H_{-\gamma}}$ is a supermartingale,

$$\mathbb{E}(B_{T\wedge H_{-\gamma}}; B_S < 0, S < H_{-\gamma}) \le \mathbb{E}(B_{S\wedge H_{-\gamma}}; B_S < 0, S < H_{-\gamma}).$$

We may rewrite the term on the left of the equation as

$$\mathbb{E}(B_T; B_S < 0, T < H_{-\gamma}) - \gamma\mathbb{P}(B_S < 0, S < H_{-\gamma} < T),$$

and by (3.10)

$$\gamma\mathbb{P}(B_S < 0, S < H_{-\gamma} < T) \le \gamma\mathbb{P}(H_{-\gamma} < T) \to 0$$

as $\gamma \to \infty$. Further, by dominated convergence,

$$\mathbb{E}(B_T; B_S < 0, T \le H_{-\gamma}) \to \mathbb{E}(B_T; B_S < 0)$$

and it follows that

$$\begin{aligned} \mathbb{E}(B_S; B_S < 0) &= \lim_{\gamma\to\infty} \mathbb{E}(B_S; B_S < 0, S < H_{-\gamma}) \\ &\ge \mathbb{E}(B_T; B_S < 0). \end{aligned}$$

Hence $\mathbb{E}B_S^- \le -\mathbb{E}(B_T; B_S < 0) \le \mathbb{E}B_T^- < \infty$.

Again using the fact that $B_{t\wedge H_{-\gamma}}$ is a supermartingale,

$$0 \ge \mathbb{E}(B_S \wedge H_{-\gamma}) = \mathbb{E}(B_S; S < H_{-\gamma}) - \gamma\mathbb{P}(H_{-\gamma} \le S)$$

so that

$$\mathbb{E}(B_S^+; S < H_{-\gamma}) \leq \mathbb{E}(B_S^-; S < H_{-\gamma}) + \gamma\mathbb{P}(H_{-\gamma} \leq S).$$

By monotone convergence the term on the left increases to $\mathbb{E}B_S^+$, while by monotone convergence and (3.10) the right hand side converges to $\mathbb{E}B_S^-$. Consequently

$$\mathbb{E}B_S^+ \leq \mathbb{E}B_S^- < \infty$$

and $\mathbb{E}|B_S| < \infty$ and $\mathbb{E}B_S \leq 0$. □

We now turn to the proof of Theorem 3.7. We will prove this theorem in several stages. We begin with a lemma whose corollary shows that the intermediate stopping time condition implies minimality. The lemma has the form given because we use this form in a later proof. For our current purposes it is the subsequent and immediate corollary which is most important.

Throughout this section it is to be understood that μ is a distribution with negative mean and T a stopping time embedding μ. Given a stopping time S let θ_S be the shift operator — the map for which $B_t(\theta_S(\omega)) = B_{S+t}(\omega)$.

Lemma 3.13. *Suppose that for all stopping times S with $S \leq T$ and $\mathbb{E}|B_S| < \infty$ we have*

$$\mathbb{E}(B_T|\mathcal{F}_S) \leq B_S \quad a.s.. \tag{3.11}$$

Then T is minimal.

Proof. Let $S \leq T$ be a stopping time such that $\mathbb{E}(B_T|\mathcal{F}_S) \leq B_S$ almost surely and such that S embeds μ (so that $\mathbb{E}|B_S| = \mathbb{E}|B_T| < \infty$). For $a \in \mathbb{R}$,

$$\begin{aligned}
\sup_{A\in\mathcal{F}_T} \mathbb{E}(a - B_T; A) &= \mathbb{E}(a - B_T; B_T \leq a) \\
&= \mathbb{E}(a - B_S; B_S \leq a) \\
&\leq \mathbb{E}(a - B_T; B_S \leq a) \qquad (3.12)\\
&\leq \sup_{A\in\mathcal{F}_T} \mathbb{E}(a - B_T; A)
\end{aligned}$$

where we use (3.11) to deduce (3.12). However since we have equality in the first and last expressions, we must also have equality throughout and so

$$\{B_T < a\} \subseteq \{B_S \leq a\} \subseteq \{B_T \leq a\}.$$

Since this holds for all $a \in \mathbb{R}$ we must have $B_T = B_S$ a.s..

Now suppose that $S \neq T$ with positive probability, and consider the stopping time $S_\epsilon = (S + H_{B_S - \epsilon} \circ \theta_S) \wedge T$. Then $\mathbb{E}|B_{S_\varepsilon}| \leq \mathbb{E}|B_S| + \mathbb{E}|B_T| + \varepsilon < \infty$. For small enough $\varepsilon > 0$, $S_\varepsilon < T$ with positive probability, and on the $\mathcal{F}_{S_\varepsilon}$-measurable set $\{B_{S_\varepsilon} = B_S - \varepsilon\}$ we have $B_{S_\varepsilon} = \mathbb{E}(B_T|\mathcal{F}_{S_\varepsilon}) - \varepsilon$ which contradicts (3.11). Consequently if (3.11) holds, $S \leq T$ and $S \sim \mu$ implies $S = T$ a.s.. □

Corollary 3.14. *Suppose that for all stopping times $S \leq T$,*

$$\mathbb{E}(B_T|\mathcal{F}_S) \leq B_S \quad a.s.. \tag{3.13}$$

Then T is minimal.

For the converse we need to show that if T is minimal then for any stopping time $S \leq T$ and $A \in \mathcal{F}_S$

$$\mathbb{E}(B_T; A) \leq \mathbb{E}(B_S; A).$$

We will use the following lemma:

Lemma 3.15. *If T is minimal then, for all $\gamma \leq 0$,*

$$f(\gamma) = \mathbb{E}(B_T - B_{T \wedge H_\gamma}) \leq 0.$$

Proof. Let $f(\gamma) = \mathbb{E}(B_T - B_{T \wedge H_\gamma}) = \mathbb{E}(B_T - \gamma; T > H_\gamma)$.

Note that $f(0) = m < 0$. Since

$$\{T \in (H_{\gamma-\varepsilon}, H_{\gamma+\varepsilon}) \setminus \{H_\gamma\}\} = \{\underline{B}_T \in (\gamma - \varepsilon, \gamma + \varepsilon), T \neq H_\gamma\}$$

and $\mathbb{P}(\underline{B}_T = \gamma, T \neq H_\gamma) = 0$, we have that

$$\mathbb{P}(T \in (H_{\gamma-\varepsilon}, H_{\gamma+\varepsilon}) \setminus \{H_\gamma\}) \leq \mathbb{P}(\underline{B}_T \in (\gamma - \varepsilon, \gamma) \cup (\gamma, \gamma + \varepsilon)).$$

Further $\mathbb{P}(\underline{B}_T \in A)$ is a probability measure on $\mathbb{R}$, so it follows from the bounded convergence theorem that

$$\mathbb{P}(T \in (H_{\gamma-\varepsilon}, H_{\gamma+\varepsilon}) \setminus \{H_\gamma\}) \to 0$$

as $\varepsilon \to 0$. The continuity of $f(\gamma)$ follows from the dominated convergence theorem and the fact that $\mathbb{E}|B_T| < \infty$, since for $0 \geq \gamma > \gamma'$ we may write

$$f(\gamma) - f(\gamma') = \mathbb{E}(B_T; H_\gamma < T < H_{\gamma'}) + (\gamma' - \gamma)\mathbb{P}(T \geq H_{\gamma'}) - \gamma\mathbb{P}(H_\gamma < T < H_{\gamma'}).$$

As a corollary if $f(\gamma_0) > 0$ for some $\gamma_0 < 0$, then there exists $\gamma_1 \in (\gamma_0, 0)$ such that $f(\gamma_1) = 0$.

Given this γ_1, and conditional on $T > H_{\gamma_1}$, let $T'' = T - H_{\gamma_1}$, $W_t = B_{H_{\gamma_1}+t} - \gamma_1$, and $\mu'' = \mathcal{L}(W_{T''})$. Suppose that T'' is not minimal, so there exists $S'' \leq T''$ with law μ''. If we define

$$S = \begin{cases} T & \text{on } T \leq H_{\gamma_1} \\ H_{\gamma_1} + S'' & \text{on } T > H_{\gamma_1} \end{cases}$$

then S embeds μ and $S \leq T$ but $S \neq T$, contradicting the minimality of T. Hence T'' is minimal. But then by Theorem 3.2, $W_{t\wedge T''}$ is uniformly integrable and so, for $\gamma < \gamma_1$

$$\mathbb{E}(W_{T''} - (\gamma - \gamma_1); T'' > H^W_{\gamma-\gamma_1}) = 0$$

or equivalently

$$f(\gamma) = \mathbb{E}(B_T - \gamma; T > H_\gamma) = 0.$$

Hence $f(\gamma) \leq 0$ for all $\gamma \in (-\infty, 0]$. $\square$

We now turn to the proof of the main result:

Proof of Theorem 3.7. We begin by showing the equivalence of conditions *(ii)* – *(v)*. It is clear that *(ii)* $\implies$ *(iii)* $\implies$ *(iv)*, the latter implication following from taking expectations in (3.8), so that when $S = T \wedge H_{-\gamma}$ we have

$$\mathbb{E}(B_T; T \leq H_{-\gamma}) - \gamma\mathbb{P}(T > H_{-\gamma}) \geq \mathbb{E}(B_T).$$

Given *(iv)* we know

$$\gamma\mathbb{P}(T > H_{-\gamma}) \leq -\mathbb{E}(B_T; T > H_{-\gamma})$$

and by dominated convergence the term on the right converges to 0 as $\gamma \to 0$ so that *(v)* holds.

For the equivalence of *(ii)* to *(v)* it only remains to show that *(v)* $\implies$ *(ii)*. So suppose *(v)* holds and choose stopping times $R \leq S \leq T$ and $A \in \mathcal{F}_R$. Set $A_\gamma = A \cap \{R < H_{-\gamma}\}$. Since $B_{t\wedge H_{-\gamma}}$ is a supermartingale

$$\mathbb{E}(B_{S\wedge H_{-\gamma}}; A_\gamma) \leq \mathbb{E}(B_{R\wedge H_{-\gamma}}; A_\gamma). \tag{3.14}$$

By Proposition 3.12 $\mathbb{E}|B_R| < \infty$ and by dominated convergence the right hand side converges to $\mathbb{E}(B_R; A)$ as $\gamma \to \infty$. For the term on the left we consider

$$\mathbb{E}(B_{S \wedge H_{-\gamma}}; A_\gamma) = \mathbb{E}(B_S; A, S < H_{-\gamma}) - \gamma \mathbb{P}(R < H_{-\gamma} < S).$$

Again by Proposition 3.12 and dominated convergence the first term on the right converges to $\mathbb{E}(B_S; A)$ while the other term converges to 0 by *(v)*. Hence on letting $\gamma \to \infty$ in (3.14) we have

$$\mathbb{E}(B_S; A) \leq \mathbb{E}(B_R; A)$$

and we have shown *(ii)*.

We have already shown that minimality is equivalent to these conditions: *(iii)* $\implies$ *(i)* is Corollary 3.14, while *(i)* $\implies$ *(iv)* is Lemma 3.15.

We have shown equivalence between *(i)* – *(v)*. We are left with showing that if μ has support bounded below then *(vi)* is also equivalent. So assume that the target distribution μ has support contained in $[\alpha, \infty)$ and that T is an embedding of μ. In that case it is easy to show that (3.9) is equivalent to (3.8). To deduce the forward implication, note that $B_{t \wedge H_\alpha}$ is a continuous supermartingale, bounded below and therefore if $S \leq T \leq H_\alpha$,

$$\mathbb{E}(B_T | \mathcal{F}_S) \leq B_S.$$

The reverse implication follows from considering the stopping time $H_{\alpha-\epsilon} = \inf\{t \geq 0 : B_t \leq \alpha - \varepsilon\}$, for then if $A = \{\omega : H_{\alpha-\epsilon} < T\}$ and $S = H_{\alpha-\varepsilon} \wedge T$,

$$(\alpha - \epsilon)\mathbb{P}(A) = \mathbb{E}(B_{H_{\alpha-\epsilon}}; A) = \mathbb{E}(B_S; A) \geq \mathbb{E}(B_T; A) \geq \alpha \mathbb{P}(A)$$

which is only possible if $\mathbb{P}(A) = 0$. □

This completes the proof of Theorem 3.7. We finish this section with a remark on an application of the theorem.

Remark 3.16. Suppose we are embedding a target distribution with negative mean m. Let T be such an embedding which consists of running the Brownian motion until it hits m, and thereafter using a (shifted) Azema-Yor embedding to embed the (zero-mean) shifted target distribution applied to $-B$, so that we maximise $\mathbb{P}(\underline{B} \leq x)$ for $x < 0$. This is an embedding we will look at more closely in Section 3.5 and has been studied by Pedersen and Peskir (2001). We show that T is minimal.

It is clear that T has the property that $\mathbb{E}(B_T; A) = m\mathbb{P}(A)$ for any set $A \in \mathcal{F}_{H_m}$, since $(B_{T \wedge (H_m + t)})_{t \geq 0}$ is a UI process.

To prove the minimality property we show that $\mathbb{E}(B_T|\mathcal{F}_S) \leq B_S$ for all stopping times $S \leq T$. Let S be such a stopping time, and suppose $A \in \mathcal{F}_S$. Then A may be written as the disjoint union, $A = (A \cap \{S \leq H_m\}) \sqcup (A \cap \{S > H_m\})$ and

$$\mathbb{E}(B_T; A) = \mathbb{E}(B_T; A \cap \{S \leq H_m\}) + \mathbb{E}(B_T; A \cap \{S > H_m\}).$$

For the first term we can deduce that

$$\mathbb{E}(B_T; A \cap \{S \leq H_m\}) = m\mathbb{P}(A, S \leq H_m) \leq \mathbb{E}(B_S; A \cap \{S \leq H_m\}).$$

For the second term, and again as a consequence of the fact that on $\{t \geq H_m\}$ the process $B_{T \wedge t}$ is UI,

$$\mathbb{E}(B_T; A \cap \{S > H_m\}) = \mathbb{E}(B_S; A \cap \{S > H_m\}).$$

Combining the results for the two terms gives

$$\mathbb{E}(B_T; A) \leq \mathbb{E}(B_S; A)$$

as required.

3.4 Constructing Minimal Stopping Times

This section is concerned with the following question:

> Suppose T is not minimal. Can we find a stopping time $T^* \leq T$ which embeds μ and is itself minimal?

The answer to this question is in the positive, as we have seen in Proposition 3.6. In this section we will demonstrate this by providing just such a construction for a given non-minimal stopping time.

The construction will be carried out in three parts. First we construct T^* in the case where μ has support on $[\alpha, \infty)$ and an atom at α. We then use limiting arguments to show that we can drop the assumption of an atom at α, and finally that we can embed μ with support on $\mathbb{R}$.

So we first assume that our support is restricted to $[\alpha, \infty)$. We will show that if T is a stopping time embedding μ, and T has the property that

$$\mathbb{P}(T > H_\alpha) > 0 \tag{3.15}$$

then we can construct a new stopping time $T^* \leq T$ which is less than H_α almost surely, and which embeds μ.

In the construction we make, we find it necessary to introduce independent randomisation. While this is undesirable, it is central to the method we use here; the randomisation will be used to 'kill' the process at rate ν. The set $\mathcal{M}^T$ we introduce below is a set of suitable 'killing measures'. In general it would seem that, for a given non-minimal stopping time, we should be able to find a non-randomised minimal stopping time which is smaller, however the specification would depend on the properties of the specific embedding.

We begin by considering the set of positive measures $\mathcal{M}^+$ on $\mathcal{B}(\mathbb{R})$ and define an ordering on $\mathcal{M}^+$ by:

$$\rho \preceq \nu \text{ iff } \rho(A) \leq \nu(A) \text{ for all } A \in \mathcal{B}(\mathbb{R}).$$

Then $\mathcal{M}^+$ is a lattice where

$$\begin{aligned} (\rho \vee \nu)(A) &= \sup\{\rho(B) + \nu(A \setminus B) : B \subseteq A\}; \\ (\rho \wedge \nu)(A) &= \inf\{\rho(B) + \nu(A \setminus B) : B \subseteq A\}. \end{aligned}$$

Also (see Doob (1984, A.IV.4)) if $\{\rho_i : i \in I\}$ is an arbitrary subset of $\mathcal{M}^+$, then there exists an order supremum of the set. To see this, we begin by assuming that the set contains every supremum of finitely many of its elements since adding these does not change the overall supremum. Define

$$\rho_*(A) = \sup_{i \in I} \rho_i(A);$$

it is clear that $\rho_* = \bigvee_{i \in I} \rho_i$ provided that ρ_* is a measure. If $A = \bigcup_{j=0}^\infty A_j$ is a countable union of disjoint measurable sets, then

$$\rho_i(A) = \sum_{j=0}^\infty \rho_i(A_j) \leq \sum_{j=0}^\infty \rho_*(A_j)$$

for all i, and hence ρ_* is countably subadditive. Finite additivity of ρ_* is trivial (given the fact that $\{\rho_i : i \in I\}$ contains the suprema of each finite set of elements) and implies

$$\rho_*(A) = \sum_{j=0}^{n} \rho_*(A_j) + \rho_* \left(\bigcup_{j=n+1}^{\infty} A_j \right) \geq \sum_{j=0}^{n} \rho_*(A_j).$$

On letting $n \to \infty$, we conclude that ρ_* is countably superadditive as well as subadditive. Hence ρ_* is a measure.

Recall that T is a stopping time that embeds μ. Given a measure $\nu \in \mathcal{M}^+$ we define the stopping time T^ν as follows.

- Let X_- and X_+ be independent random variables, independent also of B and T, and both distributed uniformly on $[0, 1]$.
- Define the levels

$$\begin{aligned} G_+^\nu &= \inf\{x \geq 0 : \nu([0, x]) \geq X_+\}, \\ G_-^\nu &= \sup\{x \leq 0 : \nu([x, 0)) \geq X_-\}, \end{aligned}$$

  and the stopping times $S_+^\nu = H_{G_+^\nu}$, $S_-^\nu = H_{G_-^\nu}$ and $S^\nu = S_+^\nu \wedge S_-^\nu$.
- Finally set

$$T^\nu = S^\nu \wedge T \wedge H_\alpha.$$

We now define the set $\mathcal{M}^T \subseteq \mathcal{M}^+$ to be

$$\mathcal{M}^T = \{\nu \in \mathcal{M}^+ : \nu((-\infty, \alpha]) = 0, \mathbb{P}\left(B_{T^\nu} \in A\right) \leq \mu(A) \ \ \forall A \in \mathcal{B}((\alpha, \infty))\}.$$

Our aim is to show that the supremum of this set, $\nu_* = \bigvee_{\nu \in \mathcal{M}^T} \nu$, is a non-zero element of $\mathcal{M}^T$, and that the stopping time $T^* \equiv T^{\nu_*}$ associated with this measure embeds μ. Since $T^* \leq H_\alpha$ it is minimal by Theorem 3.7.

Our analysis of $\mathcal{M}^T$ begins with a statement of some basic properties. Recall that we are assuming that μ has support in $[\alpha, \infty)$. Suppose also that μ has an atom at α.

Lemma 3.17. *If $\mathbb{P}(H_\alpha < T) > 0$ and $\mu(\{\alpha\}) > 0$ then the set $\mathcal{M}^T$ has the following properties:*

(i) there exists a measure $\bar{\nu} \in \mathcal{M}^+$ satisfying $\bar{\nu}(\mathbb{R}) < \infty$ and such that $\nu \in \mathcal{M}^T$ implies $\nu \preceq \bar{\nu}$;

(ii) *if* $\nu_1 \preceq \nu_2$ *then* $T^{\nu_2} \leq T^{\nu_1}$*;*

(iii) *if* $\nu_1, \nu_2 \in \mathcal{M}^T$ *then their supremum* $\nu_1 \vee \nu_2 \in \mathcal{M}^T$*;*

(iv) *if* $\nu_n \in \mathcal{M}^T$ *are a sequence of measures and* $\nu_n \uparrow \nu \in \mathcal{M}^+$ *in the sense that* $\nu_n(A)$ *is increasing for all* $A \in \mathcal{B}(\mathbb{R})$ *and*

$$\lim_n \sup_A \{|\nu_n(A) - \nu(A)| : A \in \mathcal{B}(\mathbb{R})\} = 0,$$

then $\nu \in \mathcal{M}^T$*.*

(v) *if* $\nu \in \mathcal{M}^T$ *and we define a measure* ρ *with support in* (α, ∞) *such that for* $A \in \mathcal{B}(\alpha, \infty)$*,* $\rho(A) = \mu(A) - \mathbb{P}(B_{T^\nu} \in A)$ *then* $\nu' = \nu + \rho$ *is also an element of* $\mathcal{M}^T$*.*

Proof. *(i)* Suppose $\nu \in \mathcal{M}^T$ and $x > 0$. Then $\mathbb{E}\left(B_{H_x \wedge T^\nu}\right) = 0$, so

$$\begin{aligned} |m| &= \mathbb{E}(B_{H_x \wedge T^\nu}) - \mathbb{E}(B_{T^\nu}) \\ &= \mathbb{E}(B_{H_x} - B_{T^\nu}; H_x \leq T^\nu) \\ &\leq (x - \alpha)\mathbb{P}(H_x \leq T^\nu). \end{aligned}$$

It follows that

$$\mathbb{P}\left(H_x \leq T^\nu\right) \geq \frac{|m|}{x - \alpha}. \tag{3.16}$$

In particular, for all $x > 0$, $\mathbb{P}(T^\nu \geq H_x) > 0$ and hence it must be the case that $\nu([0, x)) < 1$ for all $x > 0$.

Define a measure $\bar{\nu}$ by

$$\bar{\nu}(dx) = \begin{cases} \frac{(x-\alpha)\mu(dx)}{|m|} & x \geq 0 \\ \frac{\mu(dx)}{\mu([\alpha,x])} & \alpha < x < 0 \\ 0 & x \leq \alpha \end{cases}$$

We interpret statements about measures such as the above as shorthand for statements about the integrals over general functions. So we will write $\nu_1(dx) \leq \nu_2(dx)$ to mean that $\int f d\nu_1 \leq \int f d\nu_2$ for all positive, measurable functions f. We aim to show that $\bar{\nu}$ is an upper bound for elements of $\mathcal{M}^T$. First note that $\bar{\nu}(\mathbb{R}) < \infty$, since μ has a well defined first moment, and by assumption μ has an atom at α.

Fix $\nu \in \mathcal{M}^T$, and for $x \geq 0$ consider the probability that we stop in dx under T^ν. By definition this probability is bounded above by $\mu(dx)$. Conversely, one way to stop at

x, is to be stopped by T^ν on first reaching x. Hence

$$\mu(dx) \geq \mathbb{P}\left(H_x \leq T^\nu\right)\nu(dx) \geq \frac{|m|}{(x-\alpha)}\nu(dx),$$

and, for $x > 0$, $\nu(dx) \leq \bar{\nu}(dx)$.

If $x < 0$, since $\mathbb{P}\left(B_{T^\nu} \in A\right) \leq \mu(A)$ for all $A \in \mathcal{B}(\mathbb{R} \setminus \{\alpha\})$, we must have

$$\mathbb{P}\left(H_x \leq T^\nu\right) \geq \mu([\alpha, x]).$$

By similar arguments to above we find that $\nu \in \mathcal{M}^T$ must satisfy

$$\nu(dx) \leq \frac{\mu(dx)}{\mu([\alpha, x])} = \bar{\nu}(dx).$$

(ii) This is immediate since $\nu_1 \leq \nu_2$ implies $S^{\nu_1} \geq S^{\nu_2}$.

(iii) We wish to show that for any two measures $\nu_1, \nu_2 \in \mathcal{M}^T$, their supremum $\nu_\vee = \nu_1 \vee \nu_2$ is also in $\mathcal{M}^T$.

For $x > \alpha$ and a measure $\nu \in \mathcal{M}^T$

$$\{B_{T^\nu} \in dx\} = \{B_T \in dx, T \leq S^\nu \wedge H_\alpha\} \cup \{B_{S^\nu} \in dx, S^\nu < T \wedge H_\alpha\}. \tag{3.17}$$

If $\nu' \in \mathcal{M}^T$ and $\nu \preceq \nu'$, then $S^{\nu'} \leq S^\nu$ and

$$\{B_T \in dx, T \leq S^{\nu'} \wedge H_\alpha\} \subseteq \{B_T \in dx, T \leq S^\nu \wedge H_\alpha\}. \tag{3.18}$$

Now we consider the term $\{B_{S^\nu} \in dx, S^\nu < T \wedge H_\alpha\}$. Suppose $x > 0$ (the case $x < 0$ is similar). Fix a Brownian path $\omega = (B_t)_{0 \leq t \leq T}$, and let $\underline{B} = \underline{B}_{H_x \wedge H_\alpha} = \inf\{B_t; t \leq H_x \wedge H_\alpha\}$. The interesting case is when $T(\omega) > H_x(\omega)$. Conditional on such a path ω

$$\begin{aligned} \mathbb{P}(B_{S^\nu} \in dx, S^\nu < T \wedge H_\alpha | \omega) &= \mathbb{P}(G_+^\nu \in dx, G_-^\nu < \underline{B})\mathbf{1}_{\{\underline{B} > \alpha\}} \\ &= \nu(dx)(1 - \nu([\underline{B}, 0)))\mathbf{1}_{\{\underline{B} > \alpha\}}. \end{aligned}$$

Without loss of generality suppose $\nu_1(dx) \geq \nu_2(dx)$. Then

$$\nu_\vee(dx)(1 - \nu_\vee([\underline{B}, 0))) \leq \nu_1(dx)(1 - \nu_1([\underline{B}, 0)))$$

and

$$\mathbb{P}(B_{S^{\nu_\vee}} \in dx, S^{\nu_\vee} < T \wedge H_\alpha | \omega) \leq \mathbb{P}(B_{S^{\nu_1}} \in dx, S^{\nu_1} < T \wedge H_\alpha | \omega).$$

Combining this result with the decomposition (3.17) and the set inequality (3.18) (with $\nu' = \nu_\vee$ and $\nu = \nu_1$) we deduce

$$\mathbb{P}(B_{T^{\nu_\vee}} \in dx) \leq \mathbb{P}(B_{T^{\nu_1}} \in dx) \leq \mu(dx).$$

Hence $\nu_\vee = \nu_1 \vee \nu_2 \in \mathcal{M}^T$.

(iv) Suppose now we have a sequence of measures $\nu_n \in \mathcal{M}^T$ such that $\nu_n \uparrow \nu$. Recall the definitions of G_+^ν and G_-^ν and the fact that $\overline{\nu}(\mathbb{R}) < \infty$. Define

$$\begin{aligned} G_+^n &= \inf\{x \geq 0 : \nu_n([0,x]) \geq X_+\} \\ G_-^n &= \sup\{x \leq 0 : \nu_n([x,0)) \geq X_-\} \end{aligned}$$

where the random variables X_+ and X_- are the same random variables as those used in the definition of G_+^ν and G_-^ν. Here G_+^n is a shorthand for $G_+^{\nu_n}$. Then, for example, $\mathbb{P}(G_+^n \in A) = \nu_n(A)$ for $A \in \mathcal{B}([0,\infty))$. Consequently,

$$\mathbb{P}(G_+^n \in A) \uparrow \mathbb{P}(G_+^\nu \in A) \tag{3.19}$$

for $A \in \mathcal{B}([0,\infty))$, and similarly for G_-^ν, G_-^n. Now consider a fixed path and stopping time (so T could be determined by some independent randomisation as well as the path) of the Brownian motion, $\omega = (B_t)_{t \leq T}$, and a set $A \in \mathcal{B}(\mathbb{R})$. Then there exists a set $F = F(\omega, A) \in \mathcal{B}([0,\infty)) \times \mathcal{B}((-\infty,0))$ such that

$$B_{T^\nu} \in A \iff (G_+^\nu, G_-^\nu) \in F.$$

Note that for a fixed Brownian path, the event $B_{T^\nu} \in A$ depends on the measure ν only via the random variables G_+^ν and G_-^ν. In particular for a different measure such as ν_n we have $B_{T^{\nu_n}} \in A$ if and only if $(G_+^n, G_-^n) \in F$ for the same set $F = F(\omega, A)$.

Now

$$|\mathbb{P}(B_{T^\nu} \in A) - \mathbb{P}(B_{T^{\nu_n}} \in A)| \leq \mathbb{E}|\mathbb{P}(B_{T^\nu} \in A|\omega) - \mathbb{P}(B_{T^{\nu_n}} \in A|\omega)|.$$

Further, since $\mathbb{P}(B_{T^\nu} \in A|\omega) = \mathbb{P}((G_+^\nu, G_-^\nu) \in F(\omega))$ we have that

$$\begin{aligned} &\mathbb{P}(B_{T^\nu} \in A|\omega) - \mathbb{P}(B_{T^{\nu_n}} \in A|\omega) \\ &\qquad = \mathbb{P}((G_+^\nu, G_-^\nu) \in F(\omega)) - \mathbb{P}((G_+^n, G_-^n) \in F(\omega)) \end{aligned}$$

which tends to zero using (3.19) and its analogue for S_-. We conclude that $|\mathbb{P}(B_{T^\nu} \in A) - \mathbb{P}(B_{T^{\nu_n}} \in A)| \to 0$ and hence that $\nu \in \mathcal{M}^T$.

(v) Suppose that $\nu \in \mathcal{M}^T$ and define the measures ρ and ν' with support (α, ∞) via $\rho(A) = \mu(A) - \mathbb{P}(B_{T^\nu} \in A)$ and $\nu' = \nu + \rho$. We wish to show that ν' is also in $\mathcal{M}^T$.

As before, for $x > \alpha$ we have

$$\{B_{T_{\nu'}} \in dx\} = \{B_T \in dx, T \leq S^{\nu'} \wedge H_\alpha\} \cup \{B_{S_{\nu'}} \in dx, S^{\nu'} < T \wedge H_\alpha\},$$

and

$$\{B_T \in dx, T \leq S^{\nu'} \wedge H_\alpha\} \subseteq \{B_T \in dx, T \leq S^\nu \wedge H_\alpha\}. \tag{3.20}$$

Now suppose $x > 0$; the case $x < 0$ is similar. Conditional on a path $\omega = (B_t)_{0 \leq t \leq T}$ with the property that $H_x < T$, and with $\underline{B} = \underline{B}_{H_x \wedge H_\alpha} = \inf\{B_t; t \leq H_x \wedge H_\alpha\}$ as before,

$$\mathbb{P}(B_{S_{\nu'}} \in dx; S^{\nu'} < T \wedge H_\alpha | \omega) = \nu'(dx)(1 - \nu'([\underline{B}, 0)))\mathbf{1}_{\{(\underline{B} > \alpha)\}}.$$

It follows that

$$\begin{aligned} \mathbb{P}(B_{S_{\nu'}} \in dx; S^{\nu'} < T \wedge H_\alpha | \omega) &= (\nu(dx) + \rho(dx))(1 - \nu'([\underline{B}, 0)))\mathbf{1}_{\{(\underline{B} > \alpha)\}} \\ &\leq \nu(dx)(1 - \nu([\underline{B}, 0)))\mathbf{1}_{\{(\underline{B} > \alpha)\}} + \rho(dx) \\ &= \mathbb{P}(B_{S^\nu} \in dx; S^\nu < T \wedge H_\alpha | \omega) + \rho(dx). \end{aligned}$$

Averaging over the Brownian paths, and combining this result with (3.20) we find

$$\mathbb{P}\left(B_{T_{\nu'}} \in dx\right) \leq \mathbb{P}\left(B_{T^\nu} \in dx\right) + \rho(dx) = \mu(dx)$$

and hence $\nu' \in \mathcal{M}^T$. $\square$

We now show that T^*, our candidate for the minimal reduction of T, is an embedding. Suppose $\mathbb{P}(T > H_\alpha) > 0$, and also for the moment suppose that $\mu(\{\alpha\}) > 0$, so we may apply Lemma 3.17. The zero measure is an element of $\mathcal{M}^T$ and therefore by Lemma 3.17*(v)*, $\mathcal{M}^T$ contains a non-zero element. Now take an increasing sequence ν_i of measures such that $\nu_i \uparrow \nu_* = \bigvee \mathcal{M}^T$. We know that $\nu_* \in \mathcal{M}^T$, (Lemma 3.17*(iv)*) so the law of B_{T^*} is dominated by μ. Conversely the law of T^* also dominates μ by *(v)*. Hence T^* must embed μ.

We now note that we may drop the assumption that μ has an atom at α. Suppose μ has no atom at α. By (3.15) the law $\tilde{\mu}$ of $B_{T \wedge H_\alpha}$ does. Define

$$\mu^{(n)} = \frac{1}{n}\tilde{\mu} + \frac{n-1}{n}\mu,$$

and let $\mathcal{M}_n^T$ be the associated set of measures. Then for $k \geq n$, $\mathcal{M}_n^T \subseteq \mathcal{M}_k^T$ and so $T^{\nu_*^{(k)}} \leq T^{\nu_*^{(n)}}$ (where we write $\nu_*^{(k)}$ for the maximal element of $\mathcal{M}_k^T$). So as $n \to \infty$, $T^{\nu_*^{(n)}} \downarrow T^*$, which must therefore embed μ. Since $T^* \leq H_\alpha$, it must also be minimal.

We now consider the case of measures μ where the support is not bounded below. Define the measure μ_n by

$$\begin{aligned} \mu_n((x,y)) &= \mu((x,y)) \qquad \forall x, y \geq -n; \\ \mu_n(\{-n\}) &= \mu((-\infty,-n]); \\ \mu_n((-\infty,-n)) &= 0. \end{aligned}$$

Then for sufficiently large n, $\int x\, \mu_n(dx) < 0$. Also, $\mathcal{L}(B_{T\wedge H_{-n}})$ is dominated by both μ and μ_n on $(-n, \infty)$.

First consider the problem of embedding μ_n in $B_{T\wedge H_{-n}\wedge t}$ — that is finding a stopping time $T_n \leq T \wedge H_{-n}$ such that $\mathcal{L}(B_{T_n}) = \mu_n$. The construction above tells us that there exists a measure ν_n for which $T_n = T \wedge H_{-n} \wedge H_{G_+^n} \wedge H_{G_-^n}$ embeds μ_n, and is minimal, where

$$\begin{aligned} G_+^n &= \inf\{x \geq 0 : \nu_n([0,x]) \geq X_+\}, \\ G_-^n &= \sup\{x \leq 0 : \nu_n([x,0)) \geq X_-\}, \end{aligned}$$

for independent random variables X_+, X_- uniformly distributed on $[0,1]$.

Let n_0 be such that $\int x \mu_{n_0}(dx) < 0$. For $n_2 \geq n_1 \geq n_0$ we have that $\nu_{n_2}|_{(-n_1,\infty)} \in \mathcal{M}_{n_1}^T$ since for $A \in \mathcal{B}((-n_1,\infty))$,

$$\mathbb{P}(B_{T_{n_2}\wedge H_{-n_1}} \in A) \leq \mu_{n_2}(A) = \mu_{n_1}(A).$$

Hence for $A \in \mathcal{B}((-n_0,\infty))$, $\nu_n(A)$ decreases as n increases and we may define a measure ν_∞ by

$$\nu_\infty = \liminf_{n\to\infty} \nu_n = \sup_{k\geq 0} \inf_{n\geq k} \nu_n,$$

where the final representation ensures that ν_∞ is a measure. Our goal is to show that $T^* = S^{\nu_\infty} \wedge T$ embeds μ, and to use the fact that T_n is minimal for μ_n to deduce that T^* is minimal for μ. To this end we want to construct a coupling of the stopping times $(T_n)_{n\geq n_0}$ and T^*.

Let $\tilde{\nu}_n$ with support $(-n, \infty)$ be given by

$$\begin{aligned}
\tilde{\nu}_n([0,x]) &= \tfrac{\nu_n([0,x])-\nu_\infty([0,x])}{1-\nu_\infty([0,x])} \quad & x \geq 0; \\
\tilde{\nu}_n([x,0)) &= \tfrac{\nu_n([x,0))-\nu_\infty([x,0))}{1-\nu_\infty([x,0))} \quad & x \in (-n,0).
\end{aligned}$$

Let $\tilde{X}_+, X_+^\infty, \tilde{X}_-, X_-^\infty$ be independent random variables, independent also of B and uniformly distributed on $[0,1]$, and define the levels

$$\begin{aligned}
\tilde{G}_+^n &= \inf\{x \geq 0 : \tilde{\nu}_n([0,x]) \geq \tilde{X}_+\}, \\
G_+^\infty &= \inf\{x \geq 0 : \nu_\infty([0,x]) \geq X_+^\infty\}, \\
\tilde{G}_-^n &= \sup\{x \leq 0 : \tilde{\nu}_n([x,0)) \geq \tilde{X}_-\}, \\
G_-^\infty &= \sup\{x \leq 0 : \nu_\infty([x,0)) \geq X_-^\infty\}, \\
\bar{G}_+^n &= \tilde{G}_+^n \wedge G_+^\infty, \\
\bar{G}_-^n &= \tilde{G}_-^n \vee G_-^\infty.
\end{aligned}$$

Note that $\tilde{\nu}_n([0,x]) \downarrow 0$ as $n \uparrow \infty$ and hence $\tilde{G}_+^n \uparrow \infty$ almost surely. For $x > 0$ we have

$$\mathbb{P}(\bar{G}_+^n > x) = \mathbb{P}(\tilde{G}_+^n > x)\mathbb{P}(G_+^\infty > x) = 1 - \nu_n([x,0))$$

so that $\bar{G}_+^n$ has the same law as G_+^n defined above.

Similar calculations can be made for $x < 0$, and, using the fact that the pair $(\bar{G}_+^n, \bar{G}_-^n)$ has the same law as (G_+^n, G_-^n), we can deduce that $T_n^* = T \wedge H_{-n} \wedge H_{\bar{G}_+^n} \wedge H_{\bar{G}_-^n}$ embeds μ_n and is minimal. Furthermore, from the fact that $\tilde{G}_+^n$ and $\tilde{G}_-^n$ increase to infinity almost surely we have

$$T_n^* \uparrow T^* = T \wedge H_{G_+^\infty} \wedge H_{G_-^\infty}. \tag{3.21}$$

Since T_n^* embeds μ_n we find that T^* embeds μ.

Finally we show that T^* is minimal.

Proposition 3.18. *Suppose that T_n embeds μ_n, μ_n converges weakly to μ and $T_n \uparrow T < \infty$, almost surely. Then T embeds μ.*

If also $l_n \to l_\infty < \infty$ where $l_n = \int |x|\mu_n(dx)$ and $l_\infty = \int |x|\mu(dx)$, and T_n is minimal for μ_n, then T is minimal for μ.

Remark 3.19. The requirement that $l_n \to l_\infty$ is necessary and can be seen in the following example: consider stopping times T_n embedding $\mu_n = \frac{1}{n}\delta_{-n} + \frac{n-2}{n}\delta_0 + \frac{1}{n}\delta_n$ by running until hitting either -1 or 1, and then running until hitting either 0 or $\pm n$. Then $\mu_n \implies \delta_0$, T_n is minimal for μ_n and $T_n \uparrow T < \infty$, where T is the stopping time 'run until ± 1, then run until 0', which is not minimal.

Proof. The first part is clear, so we restrict ourselves to proving the minimality of T under the stated assumptions.

Suppose that $S \leq T$ and $A' \in \mathcal{F}_S$. By Lemma 3.13 we need to show $\mathbb{E}(B_T; A') \leq \mathbb{E}(B_S; A')$ for stopping times $S \leq T$ such that $\mathbb{E}|B_S| < \infty$. Let $A = A' \cap \{S < T\}$ and $A_n = A' \cap \{S < T_n\}$. Then

$$\mathbb{E}(B_T; A') \leq \mathbb{E}(B_S; A') \iff \mathbb{E}(B_T; A) \leq \mathbb{E}(B_S; A) \tag{3.22}$$

so we restrict our attention to sets A. Note that $A_n \uparrow A$.

Since T_n is minimal and $A_n \in \mathcal{F}_{S \wedge T_n}$

$$\mathbb{E}(B_{T_n}; A_n) \leq \mathbb{E}(B_{S \wedge T_n}; A_n) = \mathbb{E}(B_S; A_n)$$

so we deduce that both sides of (3.22) hold provided:

$$\lim_n \mathbb{E}(B_{T_n}; A_n) = \mathbb{E}(B_T; A), \tag{3.23}$$

$$\lim_n \mathbb{E}(B_S; A_n) = \mathbb{E}(B_S; A). \tag{3.24}$$

For (3.23) we consider $|\mathbb{E}(B_T; A) - \mathbb{E}(B_{T_n}; A_n)|$. Then

$$|\mathbb{E}(B_T; A) - \mathbb{E}(B_{T_n}; A_n)| \leq \mathbb{E}(|B_T|; A \setminus A_n) + \mathbb{E}(|B_T - B_{T_n}|; A_n)$$

and the first term on the right tends to zero by dominated convergence (this follows from the assumption that T_n converges to T in probability). For the second term we show $\mathbb{E}(|B_T - B_{T_n}|) \to 0$. Fix $\varepsilon > 0$. We have

$$|B_T - B_{T_n}| \leq |B_{T_n}| - |B_T| + 2|B_T|\mathbf{1}_{\{T_n \leq T - \varepsilon\}} + 2|B_T - B_{T_n}|\mathbf{1}_{\{T_n > T - \varepsilon\}}.$$

We take expectations and let $n \to \infty$. By the definition of μ_n the first two terms cancel each other out, while the third tends to zero by dominated convergence. For the last term, by the (strong) Markov property

$$\mathbb{E}(|B_T - B_{T_n}|; T_n > T - \varepsilon) \leq \mathbb{E}(|B_{T_n + \varepsilon} - B_{T_n}|) = \mathbb{E}(|B_\varepsilon|) = \sqrt{\frac{\varepsilon}{2\pi}}.$$

Consequently, in the limit, $\mathbb{E}(|B_T - B_{T_n}|; T_n > T - \varepsilon) \to 0$ and (3.23) holds.

By Lemma 3.13, we can assume that $\mathbb{E}|B_S| < \infty$ and so (3.24) follows by dominated convergence. $\square$

3.5 A Maximal Embedding for a Non-centred Target Distribution

In this section we are interested in finding an embedding to solve the following problem:

> Given a Brownian motion $(B_t)_{t\geq 0}$ and an integrable (but possibly not centred) target distribution μ with mean m, find a minimal stopping time T such that T embeds μ and
> $$\mathbb{P}(\overline{B}_T \geq x)$$
> is maximised for all x and over all minimal stopping times T embedding μ.

We call an embedding with this property the max-max embedding, and denote it by T_{max}.

Without some condition on the class of admissible stopping times the problem is clearly degenerate — any stopping time may be improved upon by waiting for the first return of the process to 0 after hitting level x and then using the original embedding. For this improved embedding $\mathbb{P}(\overline{B}_T \geq x) = 1$. Further, since no almost surely finite stopping time can satisfy

$$\mathbb{P}(\overline{B}_T \geq x) = 1$$

for all $x > 0$, there can be no solution to the problem above in the class of all embeddings. As a consequence some restriction on the class of admissible stopping times is necessary for us to have a well defined problem.

Various conditions have been proposed in the literature to restrict the class of stopping times. In the case where $m = 0$, the condition on T that $B_{t\wedge T}$ is a UI martingale has been suggested Dubins and Gilat (1978), and in this case the maximal embedding is the Azema-Yor embedding. When $m = 0$ Monroe (1972) tells us that minimality and uniform integrability are equivalent conditions, so the Azema-Yor stopping time is the max-max embedding. For the case where $m > 0$, Pedersen and Peskir (2001) showed that $\mathbb{E}\overline{B}_T < \infty$ is another suitable condition, with the optimal embedding being based on that of Azema and Yor. We argue that the class of minimal embeddings is the appropriate class for the problem under consideration since minimality is a natural and meaningful condition, which makes sense for all m (and which, for $m > 0$, includes as a subclass those embeddings with $\mathbb{E}(\overline{B}_T) < \infty$).

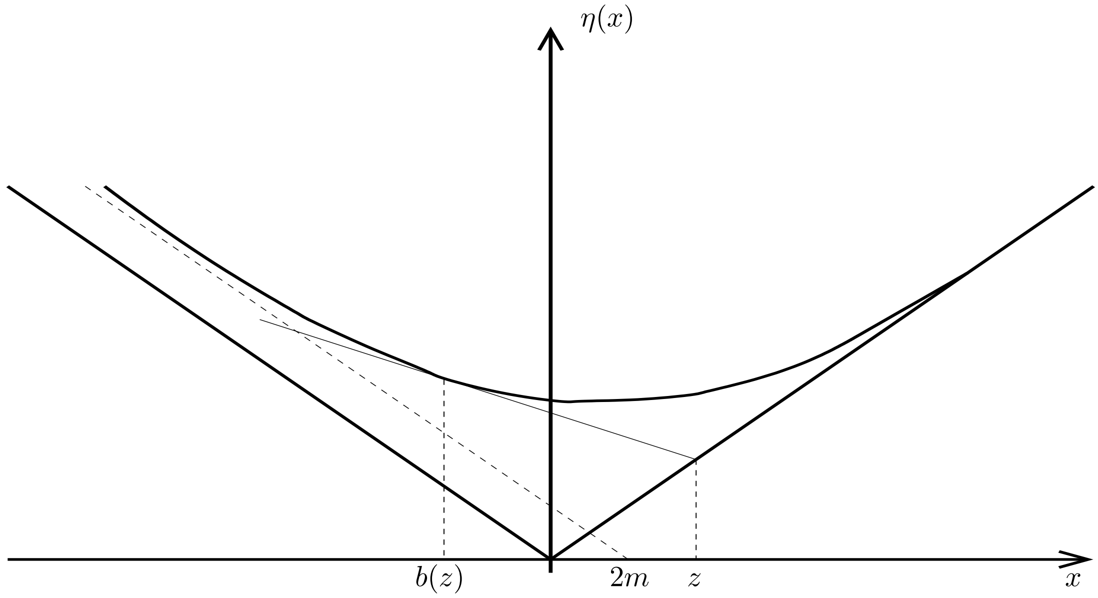

Figure 3-2: $\hat{\eta}(x)$ for a μ with support bounded above, and positive non-zero mean m. Also shown is an intuitive idea of $b(z)$.

We now describe the construction of the candidate max-max stopping time. There is some difference in the proofs of embedding and maximality between the cases where $m > 0$ and $m < 0$, however the basic idea remains the same, and much of the following construction will apply for both cases. Figures 3-2 and 3-3 show how the constructions are related.

As a refinement of (3.1), define:

$$\hat{\eta}(x) := \mathbb{E}^{\mu}|X - x| + |m|. \tag{3.25}$$

We note that as $x \to \pm\infty$, $\hat{\eta}(x) - |x| \to |m| \mp m$. The refined function $\hat{\eta}$ has the same properties as η — it is convex and Lebesgue-almost everywhere differentiable. We maintain the same definitions for u, z_+ and b, so for $\theta \in [-1, 1]$, let

$$u(\theta) := \inf\{y \in \mathbb{R} : \hat{\eta}(y) + \theta(x - y) \leq \hat{\eta}(x), \forall x \in \mathbb{R}\},$$

$$z_+(\theta) := \frac{\hat{\eta}(u(\theta)) - \theta u(\theta)}{1 - \theta},$$

and for $x \geq 0$

$$b(x) := u(z_+^{-1}(x)),$$

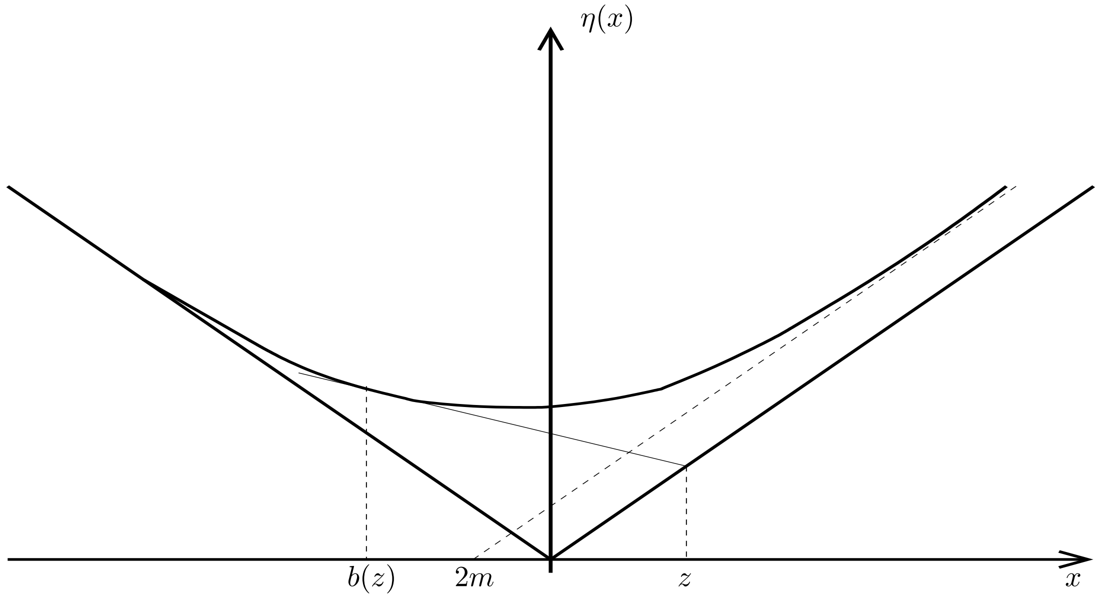

Figure 3-3: $\hat{\eta}(x)$ for a distribution μ with strictly negative mean m. Now $\hat{\eta}$ is tangential to $x + |m|$ as $x \to \infty$.

where z_+^{-1} is well defined. Finally we define the stopping time

$$T_{max} := \inf\{t > 0 : B_t \leq b(\overline{B}_t)\}. \tag{3.26}$$

As mentioned above, for $m = 0$ this is exactly the Azema-Yor stopping time, while if $m > 0$, $b(x) = -\infty$ for $x < m$, and consequently $T_{max} \geq H_m$. So when $m > 0$ this embedding may be thought of as 'wait until the process hits m then use the Azema-Yor embedding.' This is also the embedding proposed by Pedersen and Peskir (2001) and discussed in Remark 3.16. Consequently, apart from the fact that we are considering a slightly more general class of stopping times, the original part of the subsequent theorem is the case in which $m < 0$ — the rest is included for completeness.

Theorem 3.20. *Let T be a stopping time of $(B_t)_{t \geq 0}$ which embeds μ and is minimal. Then for $x \geq 0$*

$$\mathbb{P}(\overline{B}_T \geq x) \leq \left(\frac{1}{2} \inf_{\lambda < x} \frac{\hat{\eta}(\lambda) - \lambda}{x - \lambda} \right). \tag{3.27}$$

Further T_{max} embeds μ, is minimal and attains equality in (3.27) *for all $x \geq 0$.*

Remark 3.21. Note that

$$\frac{\hat{\eta}(\lambda) - \lambda}{x - \lambda} = 1 - \frac{x - \hat{\eta}(\lambda)}{x - \lambda}. \tag{3.28}$$

We can relate the right-hand-side of (3.28) to the slope of a line joining (x, x) with $(\lambda, \hat{\eta}(\lambda))$. In taking the infimum over λ we get a tangent to $\hat{\eta}$ and a value for the slope in $[-1, 1]$. Thus the bound on the right-hand-side of (3.27) lies in $[0, 1]$.

Remark 3.22. T_{max} has the property that it maximises the law of $\overline{B}_T$ over minimal stopping times which embed μ. If we want to minimise the law of the minimum, or equivalently we wish to maximise the law of $-\underline{B}_T$, then we can deduce the form of the optimal stopping time by reflecting the problem about 0, or in other words by considering $-B$. Let T_{min} be the embedding which arises in this way, so that amongst the class of minimal stopping times which embed μ, the stopping time T_{min} maximises

$$\mathbb{P}(-\underline{B}_T \geq x)$$

simultaneously for all $x \geq 0$.

The following lemma will be needed in the proof of the theorem.

Lemma 3.23. *Suppose $m \leq 0$ and T is minimal. Then for $x \geq 0$*

$$\mathbb{E}(B_{T \wedge H_x}) = 0.$$

Proof. For $x \geq 0$,

$$|B_{t \wedge T \wedge H_x}| \leq 2x - B_{t \wedge T \wedge H_x}$$

and thus

$$\mathbb{E}|B_{t \wedge T \wedge H_x}| \leq 2x - \mathbb{E}(B_{t \wedge T \wedge H_x}).$$

T is minimal so for the stopping time $S = t \wedge T \wedge H_x \leq T$, and on taking expectations in (3.8), we get

$$\mathbb{E}(B_{t \wedge T \wedge H_x}) \leq \mathbb{E}(B_T) = m.$$

Thus $\mathbb{E}|B_{t \wedge T \wedge H_x}| \leq 2x + |m|$, and by dominated convergence

$$\mathbb{E}(B_{T \wedge H_x}) = \lim_{t \to \infty} \mathbb{E}(B_{t \wedge T \wedge H_x}) = 0.$$

$\square$

We now turn to the proof of Theorem 3.20.

Proof. The following inequality for $x > 0$, $\lambda < x$ may be verified on a case by case basis:

$$\mathbf{1}_{\{\overline{B}_T \geq x\}} \leq \frac{1}{x-\lambda}\left[B_{T\wedge H_x} + \frac{|B_T-\lambda| - (B_T+\lambda)}{2}\right]. \tag{3.29}$$

In particular, on $\{\overline{B}_T < x\}$, (3.29) reduces to

$$0 \leq \begin{cases} 0 & \lambda \geq B_T \\ \frac{B_T-\lambda}{x-\lambda} & \lambda < B_T, \end{cases} \tag{3.30}$$

and on $\{\overline{B}_T \geq x\}$ we get

$$1 \leq \begin{cases} \frac{x-B_T}{x-\lambda} & \lambda > B_T \\ 1 & \lambda \leq B_T. \end{cases} \tag{3.31}$$

Then taking expectations,

$$\mathbb{P}(\overline{B}_T \geq x) \leq \frac{1}{x-\lambda}\left[\mathbb{E}(B_{T\wedge H_x}) + \frac{\hat{\eta}(\lambda) - |m| - (m+\lambda)}{2}\right]. \tag{3.32}$$

If $m \leq 0$ then by Lemma 3.23 and the minimality of T we have $\mathbb{E}(B_{T\wedge H_x}) = 0$ and so

$$\mathbb{P}(\overline{B}_T \geq x) \leq \frac{1}{2}\frac{\hat{\eta}(\lambda)-\lambda}{x-\lambda}.$$

Conversely if $m > 0$, by Theorem 3.7 applied to $-B$,

$$m = \mathbb{E}(B_T) \geq \mathbb{E}(B_{T\wedge H_x}) \tag{3.33}$$

and so

$$\mathbb{P}(\overline{B}_T \geq x) \leq \frac{1}{x-\lambda}\left[m + \frac{\hat{\eta}(\lambda) - 2m - \lambda}{2}\right] = \frac{1}{2}\frac{\hat{\eta}(\lambda)-\lambda}{x-\lambda}.$$

Since λ was arbitrary in either case, (3.27) must hold. It remains to show that T_{max} attains equality in (3.27), embeds μ and is minimal.

We begin by showing that it does attain equality in (3.27). Since

$$\frac{\hat{\eta}(\lambda)-\lambda}{x-\lambda} = 1 + \frac{\hat{\eta}(\lambda)-x}{x-\lambda}$$

the infimum in (3.27) is attained by a value λ^* with the property that a tangent of $\hat{\eta}$ at λ^* intersects the line $y = x$ at (x,x). By the definition of b we can choose $\lambda^* = b(x)$. In particular, since $\{\overline{B}_{T_{max}} < x\} \subseteq \{B_{T_{max}} \leq b(x)\}$ and $\{\overline{B}_{T_{max}} \geq x\} \subseteq \{B_{T_{max}} \geq b(x)\}$, the stopping time T_{max} attains equality almost surely in (3.30) and (3.31). Assuming that T_{max} is minimal, we are then done for $m \leq 0$. If $m > 0$ we do not always have equality in (3.33). If $x < m$ then $\mathbb{E}(B_{T_{max} \wedge H_x}) = x$, but then $\lambda^* = -\infty$ and so the extra term $(\mathbb{E}(B_{T_{max} \wedge H_x}) - m)/(x - \lambda^*) = 0$. As a result equality is again attained in (3.27). Otherwise, if $x \geq m$ then $T_{max} \geq H_m$ and the properties of the classical Azema-Yor embedding ensure that $\mathbb{E}(B_{T_{max} \wedge H_x}) = m$ and there is equality both in (3.33) and (3.27).

Fix a value of y which is less than the supremum of the support of μ, and recall that b^{-1} is defined to be left continuous. Then given (3.6) and the subsequent discussion (which remains valid even if $m \neq 0$), and since we now have equality in (3.27), we deduce:

$$\begin{aligned}
\mathbb{P}(B_{T_{max}} \geq y) &= \mathbb{P}(\overline{B}_{T_{max}} \geq b^{-1}(y)) \\
&= \frac{1}{2}\left[1 + \frac{\hat{\eta}(b(b^{-1}(y))) - b^{-1}(y)}{b^{-1}(y) - b(b^{-1}(y))}\right] \\
&= \frac{1}{2}(1 - \hat{\eta}'_-(y)) \\
&= \mu([y, \infty)).
\end{aligned}$$

Hence T_{max} embeds μ.

In the case where $m > 0$, minimality of the stopping time is discussed in Remark 3.16, while the case where $m = 0$ is discussed in Remark 3.4. We consider the case where $m < 0$. Suppose there exists $S \leq T_{max}$ such that S embeds μ. By the construction in Section 3.4 we may assume that S is minimal.

Then the following must hold:

- $B_{T_{max}} = b(\overline{B}_{T_{max}})$ by the definition of T_{max};
- $b(\overline{B}_{T_{max}}) \geq b(\overline{B}_S)$ since $T_{max} \geq S$ and $\overline{B}_{T_{max}} \geq \overline{B}_S$;
- $b(\overline{B}_S) \leq B_S$ since $S \leq T_{max} = \inf\{u : B_u \leq b(\overline{B}_u)\}$.

Hence

$$B_{T_{max}} = b(\overline{B}_{T_{max}}) \geq b(\overline{B}_S) \leq B_S.$$

If we can show $b(\overline{B}_{T_{max}}) = b(\overline{B}_S)$ a.s. then $B_{T_{max}} \leq B_S$ and since S and T_{max} embed the same law, $B_S = B_{T_{max}} = b(\overline{B}_S)$. Thus $T_{max} \leq S$ and T_{max} is minimal.

For $y \in \mathbb{R}$, consider $B_{T_{max} \wedge H_{b^{-1}(y)}}$. This random variable is distributed according to μ on $(-\infty, y)$ with a mass of size $\mu([y, \infty))$ at $b^{-1}(y)$. By the definition of the barycentre $B_{T_{max} \wedge H_{b^{-1}(y)}}$ has mean zero.

Now consider $B_{S\wedge H_{b^{-1}(y)}}$. On the set $\{H_{b^{-1}(y)} < S\}$ we have $\overline{B}_S \geq b^{-1}(y)$ and then $B_S \geq b(\overline{B}_S) \geq b(b^{-1}(y))$. Since μ assigns no mass to $(b(b^{-1}(y)), y)$ we have $B_S \geq y$. Further since $B_S \sim \mu$, then $B_{S\wedge H_{b^{-1}(y)}}$ is distributed according to μ on $(-\infty, y)$ with perhaps some mass in $[y, b^{-1}(y))$ and the rest at $b^{-1}(y)$. However S is minimal, so $\mathbb{E}(B_{S\wedge H_{b^{-1}(y)}}) = 0$ by Lemma 3.23 and in fact $B_{S\wedge H_{b^{-1}(y)}}$ assigns no mass to $[y, b^{-1}(y))$. Thus

$$\mathbb{P}(\overline{B}_S \geq b^{-1}(y)) = \mathbb{P}(\overline{B}_{T_{max}} \geq b^{-1}(y))$$

and since $\overline{B}_{T_{max}} \geq \overline{B}_S$ we conclude that the sets $\{\overline{B}_S \geq b^{-1}(y)\}$ and $\{\overline{B}_{T_{max}} \geq b^{-1}(y)\}$ are almost surely equal.

Suppose now that $\mathbb{P}(b(\overline{B}_{T_{max}}) > b(\overline{B}_S)) > 0$. Then there exists $y \in \mathbb{R}$ such that $\mathbb{P}(b(\overline{B}_{T_{max}}) > y > b(\overline{B}_S)) > 0$. But

$$\{b(\overline{B}_{T_{max}}) > y > b(\overline{B}_S)\} \subseteq \{\overline{B}_{T_{max}} > b^{-1}(y) \geq \overline{B}_S\}$$

since b^{-1} is increasing and left-continuous, and the event on the right-hand side has zero probability. $\square$

3.6 An Embedding to Maximise the Modulus

In Jacka (1988), Jacka shows how to embed a centred probability distribution in a Brownian motion so as to maximise $\mathbb{P}(\sup_{t\leq T}|B_t| \geq y)$. Our goal in this section is to extend this result to allow for non-centered target distributions with mean $m \neq 0$. In fact we solve a slightly more general problem. Let h be a measurable function; we will construct a stopping time T_{mod} which will maximise $\mathbb{P}(\sup_{t\leq T}|h(B_t)| \geq y)$ simultaneously for all y where the maximum is taken over the class of all minimal stopping times which embed μ. The reason for our generalisation will become apparent in the application in the next section.

Without loss of generality we may assume that h is a non-negative function with $h(0) = 0$ and such that for $x > 0$ both $h(x)$ and $h(-x)$ are increasing. To see this, observe that for arbitrary h we can define the function

$$\tilde{h}(x) = \begin{cases} \max_{0\leq y\leq x}|h(y)| - |h(0)| & x \geq 0; \\ \max_{x\leq y\leq 0}|h(y)| - |h(0)| & x < 0. \end{cases}$$

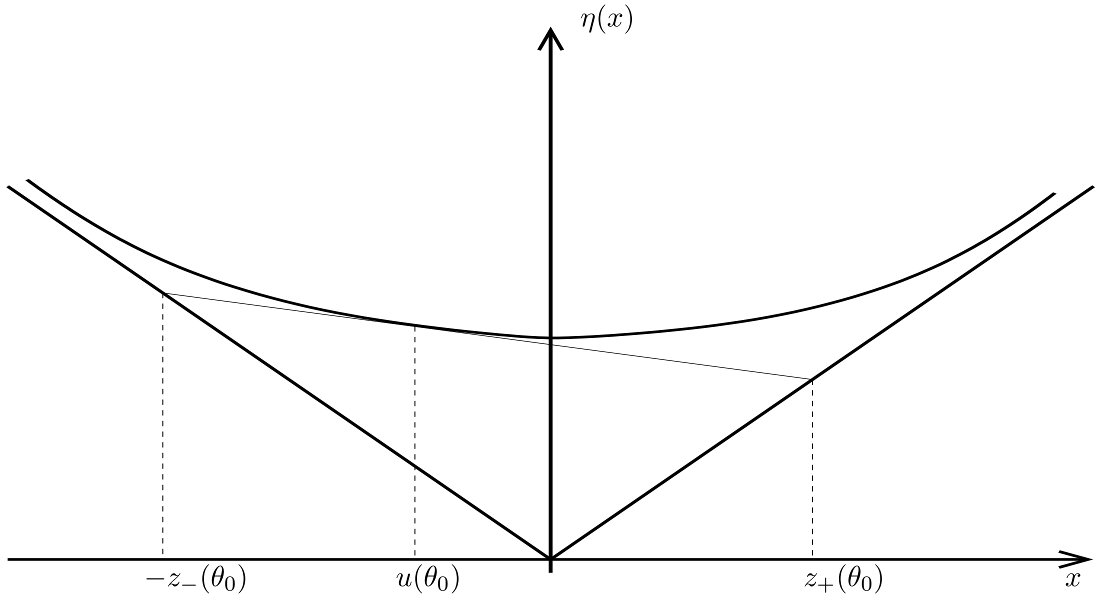

Figure 3-4: $\hat{\eta}(x)$ for a distribution μ showing the construction of $z_+(\theta_0)$ and $z_-(\theta_0)$. The slope of the tangent is θ_0 where θ_0 has been chosen such that (assuming h is continuous) $h(z_+(\theta_0)) = h(-z_-(\theta_0))$.

Then $\tilde{h}$ has the desired properties and since

$$\sup_{s \le T} |h(B_s)| = \sup_{s \le T} |\tilde{h}(B_s)| + |h(0)|$$

the optimal embedding for $\tilde{h}$ will be an optimal embedding for h.

So suppose that h has the properties listed above. We want to find an embedding of μ in B which is minimal and which maximises the law of $\sup_{t \le T} h(B_t)$. (Since h is non-negative we can drop the modulus signs.) Suppose also for definiteness that μ has a finite, positive mean $m = \int_{\mathbb{R}} x\mu(dx) > 0$. In fact our construction will also be optimal when $m = 0$ (the case covered by Jacka (1988)), but in order to avoid having to give special proofs for this case we will omit it.

We begin by making the definitions

$$z_+(\theta) := \frac{\hat{\eta}(u(\theta)) - \theta u(\theta)}{1 - \theta}, \tag{3.34}$$

$$z_-(\theta) := \frac{\hat{\eta}(u(\theta)) - \theta u(\theta)}{1 + \theta}, \tag{3.35}$$

and

$$\theta_0 := \inf\{\theta \in [-1,1] : h(z_+(\theta)) \geq h(-z_-(\theta))\},$$

as pictured in Figure 3-4. Our optimal stopping time will take the following form. Run the process until it hits either $z_+(\theta_0)$ or $-z_-(\theta_0)$, and then embed the restriction of μ to $[u(\theta_0), \infty)$ or $(-\infty, u(\theta_0)]$ respectively (defining the target measures more carefully when there is an atom at $u(\theta_0)$). For the embeddings in the second part, we will use the constructions described in Section 3.5.

To be more precise about the measures we embed in the second step, define

$$p := \mathbb{P}(H_{z_+(\theta_0)} < H_{-z_-(\theta_0)}) = \frac{z_-(\theta_0)}{z_+(\theta_0) + z_-(\theta_0)},$$

and note

$$\theta_0 = \frac{z_+(\theta_0) - z_-(\theta_0)}{z_+(\theta_0) + z_-(\theta_0)} = 1 - 2p.$$

Then let μ_+ be the measure defined by

- $\mu_+(A) = p^{-1}\mu(A)$, $A \subseteq (u(\theta_0), \infty)$, A Borel;
- $\mu_+([u(\theta_0), \infty)) = 1$;
- $\mu_+((-\infty, u(\theta_0))) = 0$,

and similarly let μ_- be given by

- $\mu_-(A) = (1-p)^{-1}\mu(A)$, $A \subseteq (-\infty, u(\theta_0))$, A Borel;
- $\mu_-((-\infty, u(\theta_0)]) = 1$;
- $\mu_-((u(\theta_0), \infty)) = 0$.

The measure μ_+ (respectively μ_-) is obtained by conditioning a random variable with law μ to lie in the upper p^{th} (respectively lower $(1-p)^{\text{th}}$) quantile of its distribution.

Recall that

$$\begin{aligned}
\hat{\eta}(y) &= \int |w - y|\mu(dw) + |m| \\
&= 2\int_{\{w>y\}} (w-y)\mu(dw) - m + y + |m|.
\end{aligned}$$

Then from the definition in (3.34) we have that

$$\begin{aligned}
z_+(\theta_0) &= \frac{1}{2p}\Bigg(2\int_{\{w>u(\theta_0)\}}(w-u(\theta_0))\,\mu(dx) - m + u(\theta_0) \\
&\qquad + |m| - (1-2p)u(\theta_0)\Bigg) \\
&= \frac{1}{p}\int_{\{w>u(\theta_0)\}} w\,\mu(dw) + u(\theta_0)\left(1 - \frac{1}{p}\int_{\{w>u(\theta_0)\}}\mu(dw)\right) \\
&\qquad + \frac{|m|-m}{2p} \\
&= \frac{1}{p}\int_{\{w>u(\theta_0)\}} w\,\mu(dw) + u(\theta_0)\left(1 - \frac{1}{p}\mu((u(\theta_0),\infty))\right)
\end{aligned}$$

where we have used $m = |m|$. In particular $z_+(\theta_0)$ is the mean of μ_+, since $\mu_+(\{u(\theta_0)\}) = 1 - \frac{1}{p}\mu((u(\theta_0),\infty))$. When we repeat the calculation for $z_-(\theta_0)$ we find that

$$\begin{aligned}
-z_-(\theta_0) =& \frac{1}{1-p}\int_{\{w<u(\theta_0)\}} w\,\mu(dw) + u(\theta_0)\left(1 - \frac{1}{1-p}\mu((-\infty,u(\theta_0))\right) \\
& - \frac{|m|+m}{2(1-p)}.
\end{aligned}$$

Since $m > 0$ the final term does not disappear and $-z_-(\theta_0)$ is strictly smaller than the mean of μ_-.

We now describe the candidate stopping time $T_{mod} \equiv T^h_{mod}$. Note that this stopping time will depend implicitly on the function h via $z_\pm(\theta_0)$. Let

$$T_0 := \inf\{t > 0 : B_t \notin (-z_-(\theta_0), z_+(\theta_0))\},$$

and define

$$T_{mod} := \begin{cases} T^{\mu_+}_{max} \circ \theta_{T_0} + T_0 & B_{T_0} = z_+(\theta_0) \\ T^{\mu_-}_{min} \circ \theta_{T_0} + T_0 & B_{T_0} = -z_-(\theta_0). \end{cases}$$

Here we use θ_{T_0} to denote the shift operator, and $T^{\mu_+}_{max}$ is the stopping time constructed in Section 3.5 for a zero-mean target distribution, so that $T^{\mu_+}_{max}$ is a standard Azema-Yor embedding of the centred target law μ_+. (Recall that $z_+(\theta_0)$ is the mean of the corresponding part of the target distribution.) Similarly $T^{\mu_-}_{min}$ is the stopping time applied to $-B$ started at $-z_-(\theta_0)$ which maximises the law of the maximum of $-B$. In this case the mean of the target law μ_- is larger than $-z_-(\theta_0)$ so that in order to define $T^{\mu-}_{min}$ we need to use the full content of Section 3.5 for embeddings of non-centred distributions.

The following theorem asserts that this embedding is indeed an embedding of μ, that it is minimal, and that it has the claimed optimality property.

Theorem 3.24. *Let μ be a target distribution such that $m > 0$. Then within the class of minimal embeddings of μ in B, the embedding T_{mod} as defined above has the property that it maximises*

$$\mathbb{P}\left(\sup_{t\le T} h(B_t) \ge x\right)$$

simultaneously for all x.

Proof. By construction T_{mod} embeds μ. We need only show that it is optimal and minimal.

Firstly, for $x \le h(-z_-(\theta_0)) \wedge h(z_+(\theta_0))$ we know that the probability of the event $\{\sup_{t\le T_{mod}} h(B_t) \ge x\}$ is one and so, for such x, T_{mod} is clearly optimal. Indeed if h is discontinuous at $-z_-(\theta_0)$ or $z_+(\theta_0)$ slightly more can be said: note first that if $z_+(\theta_0)$ coincides with the supremum of the support of μ, then by Theorem 3.7*(vi)* and the minimality of T_{mod} (see below), the stopped Brownian motion can never go above $z_+(\theta_0)$. With this in mind let

$$L = \left(\lim_{y\uparrow -z_-(\theta_0)} h(y)\right) \wedge \left(\lim_{y\downarrow z_+(\theta_0)} h(y)\right)$$

if $z_+(\theta_0)$ is less than the supremum of the support of μ and

$$L = \left(\lim_{y\uparrow -z_-(\theta_0)} h(y)\right) \wedge h(z_+(\theta_0))$$

otherwise. Now take $x \le L$. Then either $B_{T_0} = z_+(\theta_0)$ or $B_{T_0} = -z_-(\theta_0)$. If $B_{T_0} = z_+(\theta_0)$ then either $\overline{B}_{T_{mod}} > z_+(\theta_0)$ almost surely and

$$\max_{0\le t\le T_{mod}} h(B_t) \ge \lim_{y\downarrow z_+(\theta_0)} h(y) \ge L$$

or $z_+(\theta_0)$ is the supremum of the support of μ and

$$\max_{0\le t\le T_{mod}} h(B_t) = h(B_{T_0}) \ge L.$$

Similar considerations apply for $B_{T_0} = -z_-(\theta_0)$ except that then $-\underline{B}_{T_{mod}} > z_-(\theta_0)$ in all cases. We deduce that for $x \leq L$

$$\mathbb{P}\left(\sup_{t \leq T_{mod}} h(B_t) \geq x\right) = 1$$

and hence T_{mod} is optimal.

So suppose that $x > L$. For any stopping time T embedding μ, the following holds:

$$\mathbb{P}\left(\sup_{s \leq T} h(B_s) \geq x\right) \leq \mathbb{P}\left(h(\overline{B}_T) \geq x\right) + \mathbb{P}\left(h(\underline{B}_T) \geq x\right). \tag{3.36}$$

We will show that the embedding T_{mod} attains the maximal values of both terms on the right hand side, and further that for T_{mod} the two events on the right hand side are disjoint. Hence T_{mod} is optimal.

By the definition of θ_0, $x > (h(z_+(\theta_0))) \vee (h(-z_-(\theta_0)))$. It follows that

$$\mathbb{P}(h(\overline{B}_{T_{mod}}) \geq x) = p\mathbb{P}(h(\overline{B}_{T_{mod}}) \geq x | B_{T_0} = z_+(\theta_0))$$

and by the definition of T_{mod} and the properties of T_{max}, we deduce

$$\begin{aligned}\mathbb{P}(h(\overline{B}_{T_{mod}}) \geq x) &= p\mathbb{P}(h(\overline{B}_{T^{\mu}_{max}}) \geq x | \overline{B}_{T^{\mu}_{max}} \geq z_+(\theta_0)) \\ &= \mathbb{P}(h(\overline{B}_{T^{\mu}_{max}}) \geq x)\end{aligned}$$

where here T^{μ}_{max} is the embedding of Section 3.5 applied to μ. A similar calculation can be done for the minimum. In particular T_{mod} inherits its optimality property from the optimality of its constituent parts $T^{\mu_+}_{max}$ and $T^{\mu_-}_{min}$

Finally we note that T_{mod} is indeed minimal. Let $S \leq T_{mod}$ be a stopping time. We show that for $A \in \mathcal{F}_S$

$$\mathbb{E}(B_{T_{mod}}; A) \geq \mathbb{E}(B_S; A), \tag{3.37}$$

then minimality follows from Theorem 3.7. Observe that

$$B_T - B_S = B_T - B_{S \vee T_0} + B_{S \vee T_0} - B_S.$$

Then,

$$\mathbb{E}(B_T - B_{S \vee T_0}; A \cap \{T_0 = H_{z_+(\theta_0)}\}) = 0$$

by the properties of $T^{\mu_+}_{max}$;

$$\mathbb{E}(B_T - B_{S\vee T_0}; A \cap \{T_0 = H_{-z_-(\theta_0)}\}) \geq 0$$

by the properties of $T^{\mu_-}_{min}$ and

$$\mathbb{E}(B_{S\vee T_0} - B_S; A) = 0$$

since $B_{t\wedge T_0}$ is UI. If we add these various results then (3.37) follows. $\square$

Remark 3.25. If the restrictions of h to $\mathbb{R}_+$ and $\mathbb{R}_-$ are strictly increasing then T_{mod} will be essentially the unique embedding which attains optimality in Theorem 3.24. If however h has intervals of constancy then other embeddings may also maximise the law of $\sup_{t\leq T}|h(B_t)|$.

3.7 Embeddings in Diffusions

Our primary motivation in considering the embeddings of the previous sections was their use in the investigation of the following question:

> Given a regular (time-homogeneous) diffusion $(Y_t)_{t\geq 0}$ and a target distribution ν, find (if possible) a minimal stopping time which embeds ν and which maximises the law of $\sup_{t\leq T} Y_t$ (alternatively $\sup_{t\leq T}|Y_t|$) among all such stopping times.

Recall that in the martingale (or Brownian) case it is natural to consider centred target laws, at least in the first instance. However in the non-martingale case this restriction is no longer natural, and as we have seen in Chapter 2 is unrelated to whether it is possible to embed the target law in the diffusion Y.

As in Section 2.3 we use a time change to map the diffusion to a local martingale, $M_t = s(Y_t) = B_{\tau_t}$, τ being a time change which maps M to a Brownian motion. Then we can ask the question of when an embedding of Y is minimal. Clearly:

$$T \text{ is minimal for } Y \iff T \text{ is minimal for } M \iff \tau_T \text{ is minimal for } B$$

By considering Lemma 2.10, we see that when the diffusion is transient every embedding in Y is minimal, and every minimal embedding of B corresponds to an embedding of the diffusion Y. This follows from Theorem 3.7*(vi)* and Theorem 3.2.

It is now possible to apply the results of previous sections to deduce a series of corollaries about embeddings of ν in Y. Suppose that ν can be embedded in Y or equivalently that μ can be embedded in B before the Brownian motion leaves $s(I)^{\circ}$. Let T_{max} and T^h_{mod} be the optimal embeddings of μ in B as defined in Sections 3.5 and 3.6. (Observe that from now on we make the dependence of T^h_{mod} on h explicit in the notation.) Then we can define T^Y_{max} and $T^{Y,h}_{mod}$ by

$$T^Y_{max} = \tau^{-1} \circ T_{max} \qquad T^{Y,h}_{mod} = \tau^{-1} \circ T^h_{mod}.$$

Corollary 3.26. *T^Y_{max} is optimal in the class of minimal embeddings of ν in Y in the sense that it maximises*

$$\mathbb{P}\left(\max_{t \leq T} Y_t \geq y\right)$$

uniformly in $y \geq 0$.

Corollary 3.27. *$T^{Y,h}_{mod}$ is optimal in the class of minimal embeddings of ν in Y in the sense that it maximises*

$$\mathbb{P}\left(\max_{t \leq T} (h \circ s)(Y_t) \geq y\right)$$

uniformly in $y \geq 0$.

Corollary 3.28. *$T^{Y,|s^{-1}|}_{mod}$ is optimal in the class of minimal embeddings of ν in Y in the sense that it maximises*

$$\mathbb{P}\left(\max_{t \leq T} |Y_t| \geq y\right)$$

uniformly in $y \geq 0$.

Chapter 4

Extending Chacon-Walsh: Generalised Starting Distributions

In this chapter we consider a more general method for constructing stopping times, following the method of Chacon and Walsh (1976). In this context we are able to consider the embedding problem where the process has an integrable starting distribution. Consideration of general starting distributions requires a more general characterisation of minimality, and we are able to use this characterisation of minimality to provide a simple condition for the construction to be minimal.

We are then able to use the construction to provide a simple description of the stopping times considered in Chapter 3 — in particular, the Azema-Yor and modulus maximising stopping times can be easily extended to the case where we have a more general starting distribution, and shown to be minimal and optimal. Further, we are able to show that the stopping time introduced by Vallois (1983) can be described simply in the construction, which allows us to extend the stopping time to general starting measures.

4.1 Introduction

In this chapter we examine the construction of Chacon and Walsh (1976). This is essentially a graphical construction, and can be used to construct embeddings from the exit times of compact intervals. One of the features of this method is that it extends easily to generalised starting measures. The construction relies heavily on properties of the potential of the starting and target measure. Chacon and Walsh (1976) show that when the starting measure, μ_0 and the target measure μ are centred and satisfy

$$-\mathbb{E}^{\mu_0}|X-x| = u_{\mu_0}(x) \geq u_\mu(x) = -\mathbb{E}^{\mu}|X-x| \tag{4.1}$$

for all $x \in \mathbb{R}$, then the construction can be used to produce many different embeddings. One of the central aims of this chapter is to show that we can extend these results to the case where (4.1) fails, possibly because the means of the target and starting distribution are different.

In this new context we can ask the question of when the construction is minimal. When the starting distribution is simply a point mass at zero previous results concerning minimality will be relevant, however for more complicated starting distributions a new characterisation of minimality is required. This can be seen by considering the example of a target distribution consisting of a point mass at zero, but with starting distribution of mass p at -1 and $1-p$ at 1. Clearly the only minimal stopping time is to stop the first time the process hits 0, however when $p = \frac{1}{2}$ the stopping time is no longer UI, and if $p \in (0, \frac{1}{2})$ the process will not satisfy $\mathbb{E}(B_T|\mathcal{F}_S) \leq B_S$ as might be expected from consideration of Theorem 3.7. Consequently we shall require a new characterisation of minimality.

Having obtained such a characterisation, we are able to give a simple graphical interpretation of when the generalised Chacon-Walsh construction yields a minimal stopping time. These techniques allow us to demonstrate the strength of the Chacon-Walsh approach. In particular, following the techniques of Meilijson (1983), we are able to show that the Azema-Yor stopping time is a specific example of the Chacon-Walsh construction, and so we are able to construct easily an extension of the Azema-Yor stopping time for any integrable starting and target measures, showing that it does maximise the distribution of the maximum among the class of minimal stopping times.

Further we show that a second known stopping time from the literature allows an interpretation in the Chacon-Walsh picture. This time we consider the stopping time introduced in Vallois (1983), a stopping time based on the local time at zero. We are able to show that the stopping time can be easily extended to non-centred target distributions (with the process started at zero), and to some extent to more general starting distributions.

The chapter will proceed as follows: in Section 4.2 we describe the Chacon-Walsh construction, and in Section 4.3 discuss some extensions which will allow us to work more easily with non-centred distributions; in Section 4.4 we prove some technical results concerning minimality for general starting measures, which allow us in Section 4.5 to relate minimality with the potential, and prove a generalisation of Theorem 3.7; in Section 4.6 we show that (under some conditions) the limit of minimal stopping times is minimal, an observation that allows us to conclude in Section 4.7 simple sufficient conditions for a Chacon-Walsh construction to be minimal, and deduce that a generalised version of the Azema-Yor stopping time is minimal (and optimal); finally in Section 4.8 we are able to construct the Vallois stopping time in the Chacon-Walsh context, allowing us to deduce that there is a simple extension to non-centred target distributions which is minimal.

4.2 The Balayage Construction

In the theory of general Markov processes, a common definition of the potential of a stochastic process is given by

$$U\mu(x) = \int_{\mathbb{R}} \mu(dy) \int_{\mathbb{R}^+} ds\, p_s(x,y),$$

where $p_s(x,\cdot)$ is the transition density at time s of the process started at x. In the case of Brownian motion, we note that the integral is infinite. To resolve this we use the compensated definition (and introduce new notation to emphasise the fact that this is not the classical definition of potential) :

$$u_\mu(x) = \int_{\mathbb{R}} \mu(dy) \int_{\mathbb{R}^+} ds\, (p_s(x,y) - p_s(0,0)).$$

This definition simplifies to the following:

$$u_\mu(x) = -\int |x-y|\, \mu(dy). \tag{4.2}$$

If the measure μ is integrable then the function u_μ is finite for all $x \in \mathbb{R}$. It is not hard to see that (4.2) also implies that u_μ is continuous, differentiable everywhere except the set $\{x \in \mathbb{R} : \mu(\{x\}) > 0\}$ and concave.

We note that the function u_μ is connected to the functions κ of Chapter 2 and η of Chapter 3 (as in (3.1)). In particular,

$$u_\mu(x) = -\eta(x) = 2\kappa(x) - |x| - \mathbb{E}^\mu|X|. \tag{4.3}$$

The connection underlines the importance of potential theory in the study of embeddings, and the link between u_μ and η will become clear in our treatment of the Azema-Yor-type embeddings in this chapter.

In the zero-mean case, where $B_0 = 0$ and $B_T \sim \mu$ for a centred distribution μ the construction we describe is well understood. The main aim of this work is to discuss the suitability of the construction when the initial law is non-trivial and the target distribution is not centred. First we note the asymptotic behaviour of the potential. Write

$$m = \int x\, \mu(dx).$$

As $|x| \to \infty$, we have

$$u_\mu(x) + |x| \to m \operatorname{sign}(x). \tag{4.4}$$

Remark 4.1. The distribution μ is integrable if and only if $u_\mu(x)$ is finite for any (and thus all) $x \in \mathbb{R}$. It will later be important to note that, as a consequence of (4.4), if μ and ν are integrable distributions, then there exists a constant $K > 0$ such that:

$$\sup_{x \in \mathbb{R}} |u_\mu(x) - u_\nu(x)| < K.$$

Remark 4.2. The function u_μ is almost everywhere differentiable with left and right derivatives

$$\begin{aligned} u'_{\mu,-}(x) &= 1 - 2\mu((-\infty, x)); \\ u'_{\mu,+}(x) &= 1 - 2\mu((-\infty, x]). \end{aligned}$$

Chacon (1977) contains many results concerning potentials. We will describe a balayage technique that produces a sequence of measures and corresponding stopping times, and which will have as its limit our desired embedding. The following two lemmas are therefore important in concluding that the limit we obtain will indeed be the desired distribution:

Lemma 4.3 (Chacon (1977), Lemmas 2.5, 2.6)**.** *Suppose $\{\mu_n\}$ is a sequence of proba-*

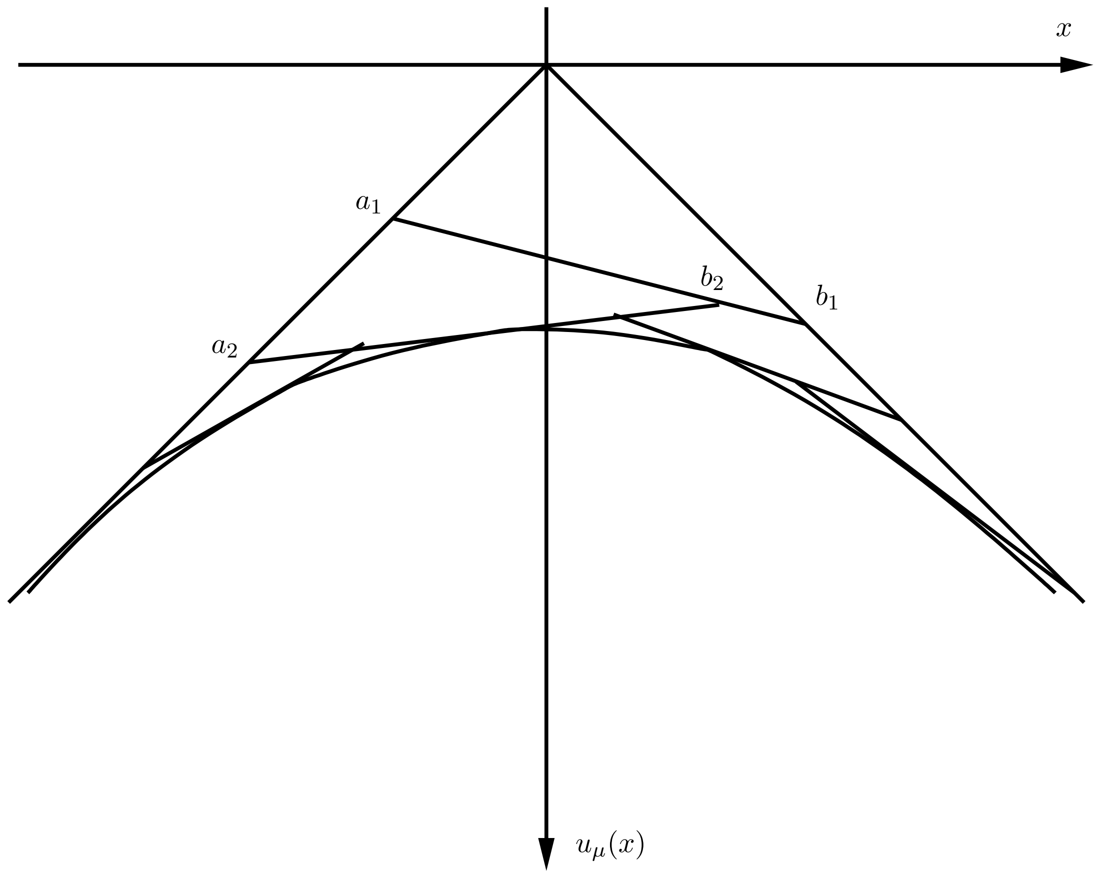

Figure 4-1: The above plot of u_μ shows some additional lines. Each line represents a step in the construction — starting from zero, we run until we hit a_1 or b_1. If we hit a_1 first, we then run until we hit a_2 or b_2 The infimum of the old potential and the line gives our new potential. In the limit we aim to have potential agreeing with the target distribution.

bility measures. If

(i) μ_n *converges weakly to* μ *and* $\lim_{n\to\infty} u_{\mu_n}(x_0)$ *exists for some* $x_0 \in \mathbb{R}$, *then* $\lim_{n\to\infty} u_{\mu_n}(x)$ *exists for all* $x \in \mathbb{R}$ *and there exists* $C \geq 0$ *such that*

$$\lim_{n\to\infty} u_{\mu_n}(x) = u_\mu(x) - C. \tag{4.5}$$

(ii) $\lim_{n\to\infty} u_{\mu_n}(x)$ *exists for all* $x \in \mathbb{R}$ *then* μ_n *converges weakly to* μ *for some measure* μ *and* μ *is uniquely determined by the limit* $\lim_n u_{\mu_n}(x)$.

We consider the embedding problem where we have a Brownian motion B with $B_0 \sim$

μ_0 (an integrable starting distribution) and we wish to embed an integrable target distribution μ. This is essentially the case considered by Chacon and Walsh (1976), although they only consider the case where $u_{\mu_0}(x) \geq u_\mu(x)$ for all x (when (4.4) implies μ_0 and μ have the same mean) — we will see that this case is simpler than the general case we consider. The embedding problem is frequently considered when μ_0 is the Dirac measure at 0. One of the appealing properties of the case where $B_0 = 0$ is that for all centred target distributions (Chacon, 1977, Lemma 2.1)

$$u_\mu(x) \leq -|x| = u_{\mu_0}(x), \tag{4.6}$$

and the condition on the ordering of potentials is easily satisfied.

One of the strengths of the Chacon-Walsh construction is that it admits a nice graphical interpretation. This is shown in Figure 4-1.

Each step in the construction is described mathematically by a simple balayage technique:

Definition 4.4. Let μ be a probability measure on $\mathbb{R}$, and I a finite, open interval, $I = (a, b)$. Then define the *balayage* μ_I of μ on I by:

$$\begin{aligned}
\mu_I(A) &= \mu(A) & A \cap \bar{I} = \emptyset; \\
\mu_I(\{a\}) &= \int_{\bar{I}} \frac{b-x}{b-a}\,\mu(dx); \\
\mu_I(\{b\}) &= \int_{\bar{I}} \frac{x-a}{b-a}\,\mu(dx); \\
\mu_I(I) &= 0.
\end{aligned}$$

The balayage μ_I is a probability measure and

$$\begin{aligned}
\int x\,\mu_I(dx) &= \int_{\bar{I}^C} x\,\mu(dx) + a\int_{\bar{I}} \frac{b-x}{b-a}\,\mu(dx) + b\int_{\bar{I}} \frac{x-a}{b-a}\,\mu(dx) \\
&= \int_{\bar{I}^C} x\,\mu(dx) + \int_{\bar{I}} x\,\mu(dx).
\end{aligned}$$

So the means of μ and μ_I agree. In particular, μ_I is the law of a Brownian motion started with distribution μ and run until the first exit from (a, b).

Our reason for introducing the Balayage technique is that the potential of μ_I is readily calculated from the potential of μ:

Lemma 4.5 (Chacon (1977) Lemma 8.1)**.** *Let μ be a probability measure with finite potential, $I = (a, b)$ a finite open interval and μ_I the balayage of μ with respect to I.*

Then

(i) $u_\mu(x) \geq u_{\mu_I}(x) \qquad x \in \mathbb{R}$;

(ii) $u_\mu(x) = u_{\mu_I}(x) \qquad x \in I^C$;

(iii) u_{μ_I} *is linear for* $x \in \bar{I}$.

Formally, we may use balayage to define an embedding as the following result shows. In the formulation of the result we assume we are given the sequence of functions we use to construct the stopping time, and from these deduce the target distribution. However we will typically use the result in situations where we have a desired target distribution and choose the sequence to fit this distribution.

Lemma 4.6. *Let* $f_1, f_2, \dots$ *be a sequence of linear functions on* $\mathbb{R}$ *such that* $|f'(x)| < 1$ *and define*

$$g(x) = \inf_{n \in \mathbb{N}} f_n(x) \wedge (u_{\mu_0}(x)). \tag{4.7}$$

Set $T_0 = 0$ *and, for* $n \geq 1$, *define*

$$\begin{aligned} a_n &= \inf\{x \in \mathbb{R} : f_n(x) < u_{\mu_{n-1}}(x)\}; \\ b_n &= \sup\{x \in \mathbb{R} : f_n(x) < u_{\mu_{n-1}}(x)\}; \\ T_n &= \inf\{t \geq T_{n-1} : B_t \notin (a_n, b_n)\}; \\ \mu_n &= (\mu_{n-1})_{(a_n, b_n)}. \end{aligned}$$

Let $T = \lim_{n\to\infty} T_n$. *If*

$$g(x) = u_\mu(x) - C \tag{4.8}$$

for some $C \in \mathbb{R}$ *and some integrable probability measure* μ *then* $T < \infty$ *a.s. and* T *is an embedding of* μ.

The condition on the gradient of the functions f_n is required to ensure that the points a_n and b_n exists.

Proof. The hard part is to show that if (4.8) holds then the stopping time T is almost surely finite. We prove in fact that $\mathbb{E}(L_T) < \infty$, where L is the local time of B at zero. By considering the martingale $|B_t| - L_t$ we must have

$$\mathbb{E}(L_{T_n}) = u_{\mu_0}(0) - u_{\mu_n}(0). \tag{4.9}$$

By monotone convergence the term on the left hand side increases to $\mathbb{E}(L_T)$ and the term on the right hand side increases to $u_{\mu_0}(0) - g(0)$, which is finite by assumption.

Lemma 4.5 implies

$$u_{\mu_n}(x) = f_n(x) \wedge u_{\mu_{n-1}}(x) = \inf_{k \leq n} f_k(x) \wedge (u_{\mu_0(x)}).$$

Since the functions f_n satisfy (4.7), we know that the conditions of Lemma 4.3 hold, determining the (unique) limiting distribution. Since $T_n \uparrow T < \infty$ a.s., B_T has distribution μ — i.e. T embeds μ — by the continuity of the Brownian motion. □

The case considered by Chacon and Walsh (1976) has a notable property. When the starting and target measures are centred (or at least when their means agree) and

$$u_{\mu_0}(x) \geq u_\mu(x) \tag{4.10}$$

then we may choose a construction such that $C = 0$ in (4.8). In this case the process $B_{t \wedge T}$ is uniformly integrable (Chacon, 1977, Lemma 5.1). The desire to find a condition to replace uniform integrability in situations where (4.10) does not hold, and to construct suitable stopping times using this framework, is the motivation behind the subsequent work.

We note also that — for given μ, μ_0 — we may choose any C which satisfies $C \geq \sup_x \{u_\mu(x) - u_{\mu_0}(x)\}$. As a consequence of (4.4) we must always have $C \geq 0$.

4.3 Non-centred Target Distributions: An Extension to Balayage

In this section we examine an extension to the method of balayage. The new step we introduce will allow us to consider a larger class of Chacon-Walsh type stopping times, and will be important when we come to consider the properties we wish our embeddings to possess, particularly in the non-zero mean case.

Definition 4.7. Let $I = (a, \infty)$ (resp. $I = (-\infty, a)$), and define the *balayage* μ_I of μ

by

$$\begin{aligned}
\mu_I(A) &= \mu(A) & A \cap \bar{I} = \emptyset; \\
\mu_I(\{a\}) &= \int_{\bar{I}} \mu(dx); & \\
\mu_I(I) &= 0. &
\end{aligned}$$

This is the distribution of a Brownian motion started with distribution μ and run until the first time it leaves I — i.e. the first time the process goes below (resp. above) the level a.

It is clear that such a stopping time is not uniformly integrable when $\mu(I) > 0$. This can be seen by noting that the means of μ_I and μ do not agree, since:

$$\int_{\mathbb{R}} x\,\mu(dx) = \int_{\mathbb{R}} x\,\mu_I(dx) + \int_I (x-a)\,\mu(dx).$$

Our aim is to classify the impact of the balayage on the potential in a similar way to Lemma 4.5.

Lemma 4.8. *Let μ be a probability measure with finite potential u_μ, $I = (-\infty, a)$ or $I = (a, \infty)$ a semi-infinite interval and μ_I the balayage of μ with respect to I. Then*

$$\begin{aligned}
u_{\mu_I}(x) &= u_\mu(x) + \Delta m & x \notin I; \\
u_{\mu_I}(x) &= u_\mu(a) + \Delta m - |a - x| & x \in I,
\end{aligned}$$

where we have written

$$\Delta m = \int_I |x - a|\,\mu(dx).$$

We may consider this graphically in the same manner as before. If we consider $u_{\mu_I} - \Delta m$ where $I = (a, \infty)$, then this function agrees with the original potential on I^C, while being a line with gradient -1 passing through the point $(a, u_\mu(a))$ on I.

The balayage step in Definition 4.7 can be recreated using the balayage steps of Definition 4.4, for example by taking the sequence of intervals $(a, a+1), (a, a+2), (a, a+3), \ldots$. However Lemma 4.5 does not tell us the resulting potential, and does not let us make the same constructions as we can with the new definition — for example if we wish our first step to be to move up to 1, we would not be able to carry out any further steps.

Lemma 4.9. *Let $f_1, f_2, \ldots$ be a sequence of linear functions on $\mathbb{R}$ such that $|f_n'(x)| \leq 1$ and*

$$g(x) = \inf_{n \in \mathbb{N}} f_n(x) \wedge (u_{\mu_0}(x)). \tag{4.11}$$

Set $T_0 = 0$, $g_0(x) = u_{\mu_0}(x)$ *and, for* $n \geq 1$, *define*

$$\begin{aligned} a_n &= \inf\{x \in \mathbb{R} : f_n(x) < g_{n-1}(x)\}; \\ b_n &= \sup\{x \in \mathbb{R} : f_n(x) < g_{n-1}(x)\}; \\ T_n &= \inf\{t \geq T_{n-1} : B_t \notin (a_n, b_n)\}; \\ g_n(x) &= g_{n-1}(x) \wedge f_n(x). \end{aligned}$$

Then the T_n *are increasing so we define* $T = \lim_{n\to\infty} T_n$. *If*

$$g(x) = u_\mu(x) - C$$

for some $C \in \mathbb{R}$ *and some integrable probability measure* μ *then* $T < \infty$ *a.s. and* T *is an embedding of* μ.

The proof of this result is essentially identical to that of Lemma 4.6, however we now have $g_n(x) = u_{\mu_n}(x) - C_n$ for some constant $C_n \in \mathbb{R}$. We also now allow $|f_n'| = 1$, which means that we might now have semi-infinite intervals. Lemma 4.8 tells us how these steps behave.

We also need to adjust the argument that $T < \infty$ a.s.. It is sufficient for this to note that e.g. $\mathbb{E}(L_{H_{-1}}) = 2$ when $B_0 = 0$, and so (4.9) holds as before. The same argument then works in this case.

The constant C chosen here is dependent on the approximating sequence of functions, but can be written as $C = u_\mu(a) - g(a)$. In the later sections we will see that the case where the functions f_n are chosen to minimise C for a given target distribution are optimal in a sense to be made explicit later.

We can now apply the graphical routine used before, along with the new 'move' introduced of drawing the line to (plus or minus) infinity with gradient plus or minus 1. This construction is shown in Figure 4-2.

The extended Chacon-Walsh embedding gives us a relatively large class of embeddings that (as a consequence of Remark 4.1) can be constructed for any integrable distributions μ, μ_0. We now turn to the question of which of these embeddings — for given starting and target distributions — are minimal.

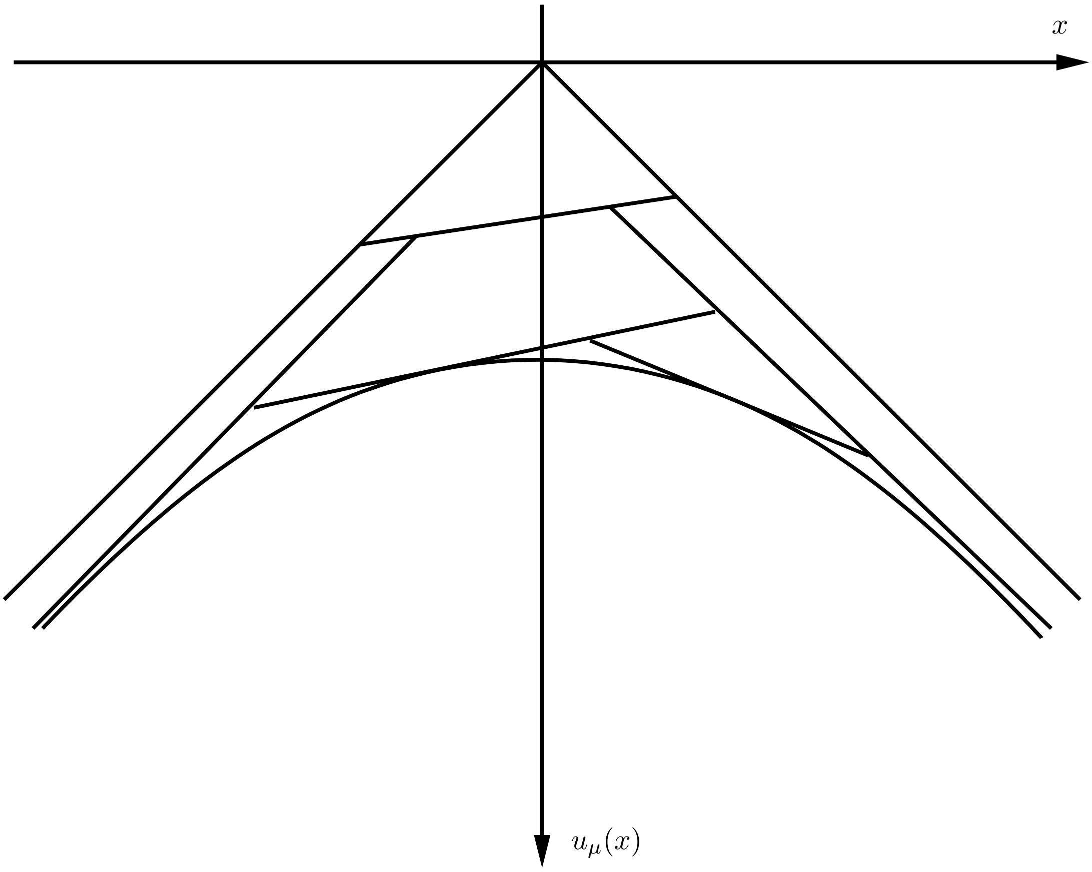

Figure 4-2: The above plot shows a potential u_μ shifted so that it is no longer tangential to $-|x|$ at either $-\infty$ or ∞. This allows steps in the construction with gradients ± 1 as shown.

4.4 Minimality: Some Preliminary Results

In this and the subsequent section we discuss necessary and sufficient conditions for an embedding of an integrable target distribution to be minimal when we have an integrable starting distribution. These results will extend the the conditions of Theorems 3.2 and 3.7. We begin by considering some of the previous results which extend easily to the general case.

As a starting point, we note that the proof of Proposition 3.6 does not rely on the fact that B starts at 0, and so the result extends to a general starting distribution, so that there always exists a minimal embedding smaller than any given embedding.

It can also be seen that the argument used in Monroe (1972) to show that if the process is uniformly integrable then the process is minimal does not require the starting measure to be a point mass. For completeness we state a similar result, with the proof identical to that given in Monroe (1972):

Lemma 4.10. *Let T be a stopping time embedding μ in $(B_t)_{t\geq 0}$, with $B_0 \sim \mu_0$ where μ and μ_0 are integrable distributions. If*

$$\mathbb{E}(B_T|\mathcal{F}_S) = B_S \text{ a.s.} \tag{4.12}$$

for all stopping times $S \leq T$ then T is minimal.

Note that $S \equiv 0$ implies that μ, μ_0 have the same mean.

Proof. Let $S \leq T$ be a stopping time such that $B_S \sim \mu$. Then for $a \in \mathbb{R}$

$$\mathbb{E}(B_T; B_T \geq a) = \mathbb{E}(B_S; B_S \geq a) = \mathbb{E}(B_T; B_S \geq a).$$

Consequently $B_S = B_T$ a.s.. If R is another stopping time, $S \leq R \leq T$, then

$$B_R = \mathbb{E}(B_T|\mathcal{F}_R) = \mathbb{E}(B_S|\mathcal{F}_R) = B_S = B_T \text{ a.s..}$$

And by the continuity of Brownian paths, B is constant on the interval $[S, T]$ and hence $S = T$ a.s.. ☐

Remark 4.11. We will later be interested also in necessary conditions for minimality. The condition in (4.12) is not necessary even when both starting and target measures are centred, as can be seen by taking $\mu_0 = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$ and $\mu = \delta_0$, where it is impossible to satisfy (4.12) but the (only) minimal stopping time is 'stop when the process hits 0.'

The condition in (4.12) is equivalent to uniform integrability of the process $(B_{t\wedge T})_{t\geq 0}$. One direction follows from the optional stopping theorem, the reverse implication comes from the upward martingale theorem (Rogers and Williams, 2000*a*)[Theorem II.69.5], which tells us that the process $X_t = \mathbb{E}(B_T|\mathcal{F}_t)$ is a uniformly integrable martingale on $t \leq T$. When (4.12) holds, $X_t = B_{t\wedge T}$, and the process $B_{t\wedge T}$ is a uniformly integrable martingale.

For the rest of this section we will consider minimality for general starting and target measures: particularly when the means do not agree. If this occurs when the starting measure is a point mass, necessary and sufficient conditions are given in Theorem 3.7. In subsequent proofs with general starting measures we will often reduce problems to the point mass case in order to apply the result.

Remark 4.12. The condition given in *(iii)* of Theorem 3.7 hints at a more general idea inherent in the study of embeddings in Brownian motion. When $B_0 = 0$, it is a well known fact that if there exists $\alpha < 0 < \beta$ such that $T \leq H_\alpha \wedge H_\beta$ then $B_{t\wedge T}$ is a uniformly integrable martingale. If $T \leq H_\alpha$ then the process is a supermartingale. In terms of embeddings, this observation has the following consequence: if the target distribution is centred and supported on a bounded interval, an embedding is minimal if and only if the process never leaves this interval. If the target distribution has a negative mean, but still lies on a bounded interval, any embedding must move above the interval — i.e. $\mathbb{P}(\sup_{t\leq T} B_t \geq x) > 0$ for all $x \geq 0$. Theorem 3.7 and Proposition 3.6 tell us that in this case an embedding exists for which $T \leq H_\alpha$ and all minimal embeddings satisfy this property.

Recall that there is a natural ordering on the set of (finite) measures on $\mathbb{R}$, that is $\mu \preceq \nu$ if and only if $\mu(A) \leq \nu(A)$ for all $A \in \mathcal{B}(\mathbb{R})$, in which case we say that ν dominates μ. In such instances it is possible to define a (positive, finite) measure $(\nu - \mu)(A) = \nu(A) - \mu(A)$. The notation $\nu = \mathcal{L}(B_T; T < H_\alpha)$ is used to mean the (sub-probability) measure ν such that $\nu(A) = \mathbb{P}(B_T \in A, T < H_\alpha)$.

Lemma 4.13. *Let B_t be a Brownian motion with $B_0 = 0$, T a stopping time embedding a distribution μ, $\tilde{\mu}$ a target distribution such that supp($\tilde{\mu}$) $\subseteq [\alpha, \infty)$ for some $\alpha < 0$ and $\int x\,\tilde{\mu}(dx) \leq 0$. Then if $\nu = \mathcal{L}(B_T; T < H_\alpha)$ is dominated by $\tilde{\mu}$, there exists a minimal stopping time $\tilde{T} \leq T \wedge H_\alpha$ which embeds $\tilde{\mu}$.*

Similarly, if $\tilde{\mu}$ is such that supp($\tilde{\mu}$) $\subseteq [\alpha, \beta]$ and $\int x\,\tilde{\mu}(dx) = 0$, and if $\nu = \mathcal{L}(B_T; T < H_\alpha \wedge H_\beta)$ is dominated by $\tilde{\mu}$, then there exists a minimal stopping time $\tilde{T} \leq T \wedge H_\alpha \wedge H_\beta$ which embeds $\tilde{\mu}$.

Proof. Construct a stopping time T' as follows: on $\{T < H_\alpha\}$, $T' = T$; otherwise choose T' so that $T' = H_\alpha + T'' \circ \theta_{H_\alpha}$ where T'' is chosen to embed $(\tilde{\mu} - \nu)$ on $\{T' \geq H_\alpha\}$ given $B_0 = \alpha$. Then T' is an embedding of $\tilde{\mu}$ and $T' \leq T$ on $\{T < H_\alpha\}$. So by Proposition 3.6 we may find a minimal embedding $\tilde{T} \leq T' \wedge H_\alpha = T \wedge H_\alpha$ which embeds $\tilde{\mu}$.

The proof in the centred case is essentially identical, but now stopping the first time the process leaves $[\alpha, \beta]$. □

We turn now to the case of interest — that is when $B_0 \sim \mu_0$ and $B_T \sim \mu$ for integrable measures μ_0 and μ. The following lemma is essentially technical in nature, but will allow us to deduce the required behaviour on letting A increase in density.

Lemma 4.14. *Let T be a minimal stopping time, and A a countable subset of $\mathbb{R}$ such that A has finitely many elements in every compact subset of $\mathbb{R}$ and $d(x, A) < M$ for all $x \in \mathbb{R}$ and some $M > 0$. We consider the stopping time*

$$R(A) = \inf\{t \geq 0 : B_t \in A\} \wedge T$$

and we write

$$E_A(x) = \begin{cases} \mathbb{E}(B_T | T > R(A), B_{R(A)} = x) & : \mathbb{P}(T > R(A), B_{R(A)} = x) > 0; \\ x & : \mathbb{P}(T > R(A), B_{R(A)} = x) = 0. \end{cases}$$

Then there exists $a \in \bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty\} \cup \{\infty\}$ such that

$$E_A(x) > x \quad \implies \quad x < a, \tag{4.13}$$

$$E_A(x) < x \quad \implies \quad x > a, \tag{4.14}$$

and $T \leq H_a$ on $\{T \geq R(A)\}$.

Further, if there exists $x < y$ such that $E_A(x) > x$ and $E_A(y) < y$ then there exists $a_\infty \in [x, y]$ such that $T \leq H_{a_\infty}$.

Proof. Suppose that there exists $x < y$ such that $E_A(x) < x$ and $E_A(y) > y$, and suppose $E_A(w) = w$ for $x < w < y$. We show that we can construct a strictly smaller embedding, contradicting the assumption that T is minimal.

Define the stopping time $T' = R(A)\mathbf{1}_{\{B_{R(A)} \in \{x,y\}\}} + T\mathbf{1}_{\{B_{R(A)} \notin \{x,y\}\}}$ and for some $z \in (x, y)$, the stopping time

$$T'' = \inf\{t \geq T' : B_t = z\} \wedge T.$$

As a consequence of Remark 4.12, paths from both x and y must hit z.

Consider the set $\{T'' < T\}$. On this set we have only paths with $B_{R(A)} = x$ and $B_{R(A)} = y$. Define $\mu_x = \mathcal{L}(B_T; B_{R(A)} = x, T'' < T)$ and $\mu_y = \mathcal{L}(B_T; B_{R(A)} = y, T'' < T)$. Since Brownian motion bounded above is a submartingale,

$$\mathbb{E}(B_{T \wedge H_z}; B_{R(A)} = x, T > R(A)) \geq x\mathbb{P}(B_{R(A)} = x, T > R(A)).$$

Together with $E_A(x) < x$ this implies

$$z\mathbb{P}(B_{R(A)} = x, T'' < T) > \mathbb{E}(B_T; B_{R(A)} = x, T'' < T),$$

i.e. we must have $\frac{1}{\mu_x(\mathbb{R})}\int w\,\mu_x(dw) < z$, and similarly $\frac{1}{\mu_y(\mathbb{R})}\int w\,\mu_y(dw) > z$. Then we apply Lemma 4.13 to the processes $B_{T''+t}$ on $\{B_{R(A)} = x, T'' < T\}$ and $\{B_{R(A)} = y, T'' < T\}$ with the measures

$$\begin{aligned}\tilde{\mu}_x &= \mu_x|_{[a_1\infty)} + \mu_y|_{(a_2,\infty)}\\ \tilde{\mu}_y &= \mu_x|_{(-\infty,a_1)} + \mu_y|_{(-\infty,a_2)}\end{aligned}$$

where we choose $a_1 < z < a_2$ so that

$$\frac{1}{\tilde{\mu}_x(\mathbb{R})}\int w\,\tilde{\mu}_x(dw) \le z \text{ and } \frac{1}{\tilde{\mu}_y(\mathbb{R})}\int w\,\tilde{\mu}_y(dw) \ge z$$

and also so that $\mu_x(\mathbb{R}) = \tilde{\mu}_x(\mathbb{R})$ and $\mu_y(\mathbb{R}) = \tilde{\mu}_y(\mathbb{R})$[1]. This will produce a strictly smaller embedding, in contradiction to the assumption that T is minimal.

So we have shown that there exists a such that (4.13) and (4.14) hold. We just need to show that we can choose a so that $T \le H_a$ on $\{T \ge R(A)\}$.

Suppose that there exists $x < y$ such that $E_A(x) > x$ and $E_A(y) < y$ and $E_A(w) = w$ for $w \in (x, y)$. If

$$\sup_{x<a} \mu_x((a,\infty)) = 0 \text{ and } \sup_{y>a} \mu_y((-\infty,a)) = 0 \text{ for some } a \in (x,y) \tag{4.15}$$

then T minimal and Theorem 3.7 implies that $T \le H_a$ on $\{T \ge R(A)\}$.

So suppose that (4.15) does not hold. We shall show that we can find a sequence $x_1, x_2, \ldots, x_r$ of elements of A such that we are able to transfer mass between the x_i to produce a smaller embedding. We begin by choosing x_1 to be the point of A satisfying $E_A(x) > x$ for which the support of μ_x extends furthest to the right, and y_1 similarly the point satisfying $E_A(y) < y$ for which the support of μ_y extends furthest to the left. If the support of these measures overlap we show we can exchange mass between μ_{x_1} and μ_{y_1} and embed to find a smaller stopping time. Otherwise we look at those points for which $E_A(x) = x$ and the support overlaps that of μ_{x_1} but extends further to the right. In this way we can find a sequence whose supports overlap (since (4.15) does not hold) and we may again perform a suitable exchange of mass to show that we can find a smaller embedding. Then we take $x_r = y_1$ and the points satisfy $x_2 < x_3 < \ldots < x_{r-1}$.

There are several technical issues we need to address. Firstly, if we find at some stage

[1] It may be necessary to consider only a proportion of the paths hitting z from one side; this can be done by choosing paths according to an independent $U([0,1])$ random variable and running the rest of the paths according to T. This will still construct a strictly smaller stopping time.

there are two points which both satisfy the criterion — for example their supports have the same upper bound — then we may use either point. Secondly, if the support of all suitable points has a maximum which is not attained we may still use the same procedure but we must (and can) choose a point which approximates the bound suitably closely for subsequent steps to work. Finally we note that once we choose x_2, since there is at most one point to the right of y_1, there exists only a finite number of points left to choose from (by assumption on A) and so the sequence will be finite.

The technical construction is as follows: let x_1 be the largest value such that $E_A(x_1) > x_1$ and

$$\sup\{z : z \in \mathrm{supp}(\mu_{x_1})\} = \sup_{w:E_A(w)>w} \{\sup\{z : z \in \mathrm{supp}(\mu_w)\}\},$$

(or at least so that the left hand side approximates the right hand side sufficiently closely for the next step to work — since the support of the points to the right overlaps we shall be able to find x_1 with supremum of its support sufficiently close to the term on the left) and let y_1 be the smallest value such that $E_A(y_1) < y_1$ and

$$\inf\{z : z \in \mathrm{supp}(\mu_{y_1})\} = \inf_{w:E_A(w)<w} \{\inf\{z : z \in \mathrm{supp}(\mu_w)\}\}.$$

Then (by the assumption that (4.15) does not hold) we can find a sequence $x_1, x_2, \ldots, x_r$ such that $x_r = y_1$ and $x_2 < x_3 < \ldots x_{r-1}$, $E_A(x_i) = x_i$ for $1 < i < r$ and, if we define $I_i = \inf\{\text{intervals } I : \mathrm{supp}(\mu_{x_i}) \subseteq I\}$, then

$$\begin{aligned} Leb(I_i \cap I_{i+1}) &> 0 \qquad\qquad k = 1, \ldots, r-1, \\ Leb(I_i \cap I_{i+2}) &= 0 \qquad\qquad k = 1, \ldots, r-2. \end{aligned}$$

This is done by choosing at each step the w with $E_A(w) = w$ which overlaps the support of the previous μ_{x_i} and whose support extends furthest to the right, until the support overlaps with the support of μ_{y_1}.

We write $\mu_i = \mu_{x_i}$. For general $1 \leq i < r$ now consider μ_i' defined by

$$\mu_i' = \mu_i|_{(-\infty,y_i)} + \mu_{i+1}|_{(-\infty,y_i)}$$

where y_i is chosen such that $\mu_i([y_i, \infty)) = \mu_{i+1}((-\infty, y_i))$. Then it must be true that

$\int w\,\mu_i'(dw) < \int w\,\mu_i(dw)$. Define

$$\begin{aligned} m_i &= \int w\,\mu_i(dw) - \int w\,\mu_i'(dw) > 0 \\ m_0 &= \int w\,\mu_1(dw) - \mu_1(\mathbb{R})x_1 > 0 \\ m_r &= \mu_r(\mathbb{R})x_r - \int w\,\mu_r(dw) > 0 \end{aligned}$$

and set $\Delta m = \inf\{m_i : 0 \leq i \leq r\}$. Then for each i we can find $v_i < z_i$ such that $\mu_i([z_i, \infty)) = \mu_{i+1}((-\infty, v_i))$ and for

$$\mu_i' = \mu_i|_{(-\infty,z_i)} + \mu_{i+1}|_{(-\infty,v_i)}$$

we have

$$\int w\,\mu_i(dw) - \int w\,\mu_i'(dw) = \Delta m.$$

Set

$$\begin{aligned} \mu_1'' &= \mu_1|_{(-\infty,z_1)} + \mu_2|_{(-\infty,v_1)}, \\ \mu_i'' &= \mu_{i-1}|_{[z_{i-1},\infty)} + \mu_i|_{[v_{i-1},z_i)} + \mu_{i+1}|_{(-\infty,v_i)} \qquad\qquad i = 2, \ldots, r-1, \\ \mu_r'' &= \mu_{r-1}|_{[z_{r-1},\infty))} + \mu_r|_{[v_{r-1},\infty)}. \end{aligned}$$

Then

$$\begin{aligned} \int x\,\mu_1'' &\geq \mu_1''(\mathbb{R})x_1 \\ \int x\,\mu_i'' &= \mu_i''(\mathbb{R})x_i \qquad\qquad i = 2, \ldots, r-1 \\ \int x\,\mu_r'' &\leq \mu_r''(\mathbb{R})x_r. \end{aligned}$$

So the conditions of Lemma 4.13 are satisfied for each μ_i'' and we can find strictly smaller stopping times on each of the sets $\{T > R(A), R(A) = x_i\}$.

It only remains to show the final statement of the lemma. Let $A' \supset A$ be another set satisfying the conditions of the lemma for some M', such that $A' \setminus A \subseteq [x, y]$. Then there exists $x', y' \in A'$ such that $x \leq x' < y' \leq y$, $E_{A'}(x') > x'$ and $E_{A'}(y') < y'$ — if this were not the case at least one of the embeddings conditional on $\{R(A') = z\}$ would not be minimal.

Now consider a sequence $A \subset A_1 \subset A_2 \subset \ldots$ and such that $A_n \setminus A \subseteq [x, y]$ and

$d(z, A_n) \leq 2^{-n}$ for $z \in [x, y]$. Let

$$\begin{aligned} \Lambda &= \{a \in [x,y] : T \leq H_a \text{ on } \{T \geq R(A)\}\}; \\ \Lambda_n &= \{a \in [x,y] : T \leq H_a \text{ on } \{T \geq R(A_n)\}\}. \end{aligned}$$

Then the sets Λ, Λ_n are closed, $\Lambda \supseteq \Lambda_1 \supseteq \Lambda_2 \supseteq \ldots$, and each Λ_n is non-empty. So there exists $a_\infty \in \Lambda_n$ for all n. Hence $T \leq H_{a_\infty}$ on $\{T \geq R(A_n)\}$ for all n. But $R(A_n) \downarrow 0$ on $\{B_0 \in [x, y]\}$ and $R(A) \leq H_{a_\infty}$ on $\{B_0 \notin [x, y]\}$. □

This result, although technical in nature, can be thought of as beginning to describe the sort of behaviour we shall expect from minimal embeddings in this general context. The cases considered in Chapter 3 suggest behaviour of the form: 'the process always drifts in the same direction', if indeed it drifts at all. The example of Remark 4.11 suggests that this is not always possible in the general case, and the previous result suggests that this is modified by breaking the space into two sections, in each of which the process can be viewed separately. The way these sections are determined is clearly dependent on the starting and target measures, and we shall see in the next section that the potential of these measures provides an important tool in determining how this occurs.

4.5 Minimality and Potential

The main aim of this section is to find equivalent conditions to minimality which allow us to characterise minimality simply in terms of properties of the process $B_{t \wedge T}$. This is partly in order to prove the following result:

> The Chacon-Walsh type embedding is minimal when constructed using the functions u_{μ_0} and $c(x) = u_\mu(x) - C$ where
>
> $$C = \sup_x \{u_\mu(x) - u_{\mu_0}(x)\}. \tag{4.16}$$

We have already shown that provided the means of our starting and target distribution match, and (4.10) holds (so that $C = 0$ — the solution in this case to (4.16)), then the process constructed using the Chacon-Walsh technique is uniformly integrable, and therefore minimal. Of course the Chacon-Walsh construction is simply an example of an embedding, and the functions u_{μ_0} and c are properties solely of the general problem

— it seems reasonable however that these functions will appear in the general problem of classifying all minimal embeddings.

So consider a pair μ_0, μ of integrable measures. Remark 4.1 tells us we we can choose C such that (4.16) holds. We know $u_{\mu_0}(x) - c(x)$ is bounded above, and $\inf_{x\in\mathbb{R}} u_{\mu_0}(x) - c(x) = 0$. We consider

$$\mathcal{A} = \{x \in [-\infty, \infty] : \lim_{y\to x} u_{\mu_0}(y) - c(y) = 0\}. \tag{4.17}$$

Since both functions are Lebesgue almost-everywhere differentiable, Remark 4.2 implies $\mathcal{A} \subseteq \mathcal{A}'$ where $\mathcal{A}'$ is the set

$$\{x \in [-\infty, \infty] : \mu((-\infty, x)) \leq \mu_0((-\infty, x)) \leq \mu_0((-\infty, x]) \leq \mu((-\infty, x])\}. \tag{4.18}$$

One consequence of this is that if the starting distribution has an atom at a point of $\mathcal{A}$ then the target distribution has an atom at least as large. Also we introduce the following definition. Given a measure ν, $a \in \mathbb{R}$ and $\theta \in [\nu((-\infty, a)), \nu((-\infty, a])]$ we define the measure $\check{\nu}^{a,\theta}$ to be the measure which is ν on $(-\infty, a)$, has support on $(-\infty, a]$ and $\check{\nu}^{a,\theta}(\mathbb{R}) = \theta$. We also define $\hat{\nu}^{a,\theta} = \nu - \check{\nu}^{a,\theta}$. Then for $a \in \mathcal{A}$ we may find θ such that

$$\begin{aligned}\check{\mu}^{a,\theta}((-\infty, a]) &= \check{\mu}_0^{a,\theta}((-\infty, a]) \\ \hat{\mu}^{a,\theta}([a, \infty)) &= \hat{\mu}_0^{a,\theta}([a, \infty)).\end{aligned}$$

When $\mu_0((-\infty, a)) < \mu_0((-\infty, a])$ there will exist multiple θ. We will occasionally drop the θ from the notation since this is often unnecessary.

These definitions allows us to write the potential in terms of the new measures (for any suitable θ)

$$u_\mu(x) = \int_{(-\infty,x]} (y - x)\, \check{\mu}^x(dy) + \int_{[x,\infty)} (x - y)\hat{\mu}^x(dy). \tag{4.19}$$

As a consequence of this and a similar relation for u_{μ_0}, we are able to deduce the following important facts about the set $\mathcal{A}$:

- if $x < z$ are both elements of $\mathcal{A}$ (possibly $\pm\infty$), then

$$\int y\,(\mu - \check{\mu}^{x,\theta} - \hat{\mu}^{z,\phi})(dy) = \int y\,(\mu_0 - \check{\mu}_0^{x,\theta} - \hat{\mu}_0^{z,\phi})(dy). \tag{4.20}$$

  That is, we may find measures agreeing with μ and μ_0 on (x, z) and with support on $[x, z]$ which have the same mean.

- If $x \in \mathcal{A}$, by definition

$$u_\mu(x) - u_{\mu_0}(x) \geq \lim_{z \to -\infty} (u_\mu(z) - u_{\mu_0}(z)). \tag{4.21}$$

This can be rearranged, using (4.19), to deduce

$$\int_{(-\infty,x]} y\, \check{\mu}_0^x(dy) \leq \int_{(-\infty,x]} y\, \check{\mu}^x(dy)$$

with equality if and only if there is also equality in (4.21) — that is when $-\infty \in \mathcal{A}$.

Together these imply that the set $\mathcal{A}$ divides $\mathbb{R}$ into intervals on which the starting and target measures place the same amount of mass. Further, the means of the distributions agree on these intervals except for the first (resp. last) interval where the mean of the target distribution will be larger (resp. smaller) than that of the starting distribution unless $-\infty$ (resp. ∞) is in $\mathcal{A}$, when again they will agree. Note the connection between this idea and Lemma 4.14

Before we prove the result we establish several results that are needed in the proof.

Proposition 4.15. *Suppose $T \leq H_{a_\infty}$ is an embedding of μ for $a_\infty \in \mathbb{R}$. Then $a_\infty \in \mathcal{A}$.*

Proof. Clearly a_∞ must lie in $\mathcal{A}'$ (see (4.18)). Suppose also that $a_\infty < z \in \mathcal{A}$. We may choose θ, ϕ such that $\mu_0 - \check{\mu}_0^{a_\infty,\theta} - \hat{\mu}_0^{z,\phi}$ has no atom at either a_∞ or z.

Then

$$u_{\mu_0}(a_\infty) \geq u_\mu(a_\infty) - C \tag{4.22}$$

and $C = u_\mu(z) - u_{\mu_0}(z)$ imply

$$\int y\,(\mu - \check{\mu}^{a_\infty,\theta} - \hat{\mu}^{z,\phi})(dy) \geq \int y\,(\mu_0 - \check{\mu}_0^{a_\infty,\theta} - \hat{\mu}_0^{z,\phi})(dy), \tag{4.23}$$

the term on the right being equal to $\mathbb{E}(B_0; B_0 \in (a_\infty, z))$ and the term on the left at most $\mathbb{E}(B_T; B_0 \in (a_\infty, z))$. However $B_T = B_{T \wedge H_{a_\infty}}$ is a supermartingale on $\{B_0 \geq a_\infty\}$, so we must have equality in (4.23) and hence in (4.22). So $a_\infty \in \mathcal{A}$. ☐

Proposition 4.16. *Suppose T is minimal and A is a countable subset of $\mathbb{R}$ such that A has finitely many elements in every compact subset of $\mathbb{R}$ and $d(x, A) < M$ for all $x \in \mathbb{R}$ and some $M > 0$. Suppose also that $S \leq T$ is a stopping time and $I \subseteq \mathbb{R}$ is an interval such that $\partial I \subseteq A$. If*

$$\mathbb{E}(B_T; F \cap \{B_0 \in I\}) > \mathbb{E}(B_S; F \cap \{B_0 \in I\}) \tag{4.24}$$

for some $F \in \mathcal{F}_S$ then $E_A(x) > x$ for some $x \in A \cap \bar{I}$.

Proof. We may assume $F \subseteq \{B_0 \in I\}$ and we note that therefore $B_{R(A)} \in A \cap \bar{I}$ on $\{R(A) < T\} \cap F$. Since $B_{t \wedge R(A)}$ is uniformly integrable,

$$\begin{aligned}
\mathbb{E}(B_S; F) &= \mathbb{E}(B_{R(A)}; F \cap \{S \leq R(A)\}) + \mathbb{E}(B_S; F \cap \{R(A) < S\}) \\
\mathbb{E}(B_T; F) &= \mathbb{E}(B_{R(A)}; F \cap \{T = R(A)\}) + \mathbb{E}(B_T; F \cap \{R(A) < T\}).
\end{aligned}$$

So (4.24) and the above identities imply

$$\begin{aligned}
\mathbb{E}(B_T; F \cap \{R(A) < T\}) >&\mathbb{E}(B_{R(A)}; F \cap \{S \leq R(A) < T\}) \\
&+ \mathbb{E}(B_S; F \cap \{R(A) < S\}).
\end{aligned}$$

However if $E_A(x) \leq x$ for all $x \in A \cap \bar{I}$ and T is minimal, by Theorem 3.7:

$$\begin{aligned}
\mathbb{E}(B_T; F \cap \{R(A) < S\}) &\leq \mathbb{E}(B_S; F \cap \{R(A) < S\}) \\
\mathbb{E}(B_T; F \cap \{S \leq R(A) < T\}) &\leq \mathbb{E}(B_{R(A)}; F \cap \{S \leq R(A) < T\})
\end{aligned}$$

and we deduce a contradiction. □

Proposition 4.17. *Suppose $F \in \mathcal{F}_0$, $\mathbb{E}(B_T; F) = \mathbb{E}(B_0; F)$ and*

$$\mathbb{E}(B_T | \mathcal{F}_S) \leq B_S \text{ on } F \tag{4.25}$$

for all stopping times S. Then in fact we have equality — that is

$$\mathbb{E}(B_T | \mathcal{F}_S) = B_S$$

almost surely on F.

Proof. If $\mathbb{P}(F) = 0$ there is nothing to prove. Otherwise we may condition on F to reduce to showing the result when $F = \Omega$.

By the upward martingale theorem (Rogers and Williams, 2000*a*)[Theorem II.69.5], the process

$$X_t = \mathbb{E}(B_T | \mathcal{F}_t)$$

is uniformly integrable. Also $\mathbb{E}(B_T | \mathcal{F}_0) \leq B_0$ and $\mathbb{E}(B_T) = \mathbb{E}(B_0)$ implies $\mathbb{E}(B_T | \mathcal{F}_0) = B_0$. Let $Y_t = B_{T \wedge t} - X_{T \wedge t}$. By (4.25) Y_t is a non-negative local martingale such that $Y_0 = Y_T = 0$. Hence $Y \equiv 0$. □

Lemma 4.18. *If T is minimal and $a \in \mathcal{A}$ then $T \leq H_a$ and*

$$\begin{aligned} \mathbb{E}(B_T|\mathcal{F}_S) &\leq B_S \text{ on } \{B_0 \geq a\}; & (4.26)\\ \mathbb{E}(B_T|\mathcal{F}_S) &\geq B_S \text{ on } \{B_0 \leq a\}. & (4.27) \end{aligned}$$

Proof. Suppose initially $a \in \mathbb{R}$. Let $\theta = \mu_0((-\infty, a))$. If $\{B_0 < a\} \not\subseteq \{B_T \leq a\}$ a.s. then also $\{B_0 \geq a\} \not\subseteq \{B_T \geq a\}$ a.s. and

$$\begin{aligned} \mathbb{E}(B_0; B_0 < a) = \int y\, \check{\mu}_0^{a,\theta}(dy) &\leq \int y\, \check{\mu}^{a,\theta}(dy) < \mathbb{E}(B_T; B_0 < a);\\ \mathbb{E}(B_0; B_0 \geq a) = \int y\, \hat{\mu}_0^{a,\theta}(dy) &\geq \int y\, \check{\mu}^{a,\theta}(dy) > \mathbb{E}(B_T; B_0 < a). \end{aligned}$$

So there exists $x_1 \leq a$ and $x_2 \geq a$ such that (by Proposition 4.16)

$$E_A(x_1) < x_1 \text{ and } E_A(x_2) > x_2$$

for a suitable choice of A — a contradiction to Lemma 4.14.

A similar argument can be used with $\theta = \mu_0((-\infty, a])$ to deduce that $\{B_0 \leq a\} \subseteq \{B_T \leq a\}$ a.s. and $\{B_0 \geq a\} \subseteq \{B_T \geq a\}$ a.s.. So if there is an atom of μ_0 at a then paths starting at a must also stop at a, and hence (by the minimality of T) must stop immediately — i.e. $T = 0$ on $\{B_0 = a\}$.

So consider paths for which $\{B_0 < a\}$. For almost all these paths, for some choice of A, $B_{R(A)} < a$. If (4.27) fails, by Proposition 4.16 there exists $x < a$ such that $E_A(x) < x$. Then Lemma 4.14 and (for $\theta = \mu_0((-\infty, a))$)

$$\int y\, \check{\mu}_0^{a,\theta}(dy) \leq \int y\, \check{\mu}^{a,\theta}(dy)$$

imply there must also exist $y < x$ such that $E_A(y) > y$, and hence $a' < a$ such that $T \leq H_{a'}$. Then $B_{t\wedge T}$ is a supermartingale on $\{B_0 > a'\}$ (and a submartingale on $\{B_0 \leq a'\}$). But Proposition 4.15 and (4.20) imply $\mathbb{E}(B_0; a' < B_0 < a) = \mathbb{E}(B_T; a' < B_0 < a)$ and therefore (by Proposition 4.17) $B_{t\wedge T}$ is a true martingale on $\{a' < B_0 < a\}$ — in particular $T \leq H_a$ on $\{B_0 < a\}$, and (4.27) holds. Similarly (4.26) can be shown to hold.

So suppose now that $a = \infty$ (the case $a = -\infty$ is similar) and there exists $a' < \infty$ also in $\mathcal{A}$. By the above, $T \leq H_{a'}$ and so $B_{t\wedge T}$ is a supermartingale on $\{B_0 > a'\}$, while by (4.20) $\mathbb{E}(B_0; B_0 > a') = \mathbb{E}(B_T; B_0 > a')$, and hence $B_{t\wedge T}$ satisfies (4.27) by Proposition 4.17.

Finally suppose $\mathcal{A} = \{\infty\}$. By Lemma 4.14 $E_A(x) \geq x$ for all suitable choices of A and all x. Hence, by Proposition 4.16,

$$\mathbb{E}(B_T|\mathcal{F}_S) \geq B_S.$$

$\square$

We note that some of the above arguments, particularly the use of Proposition 4.17, allow us to deduce that if there exists $a \in \mathcal{A}$, $|a| < \infty$ for which $T \leq H_a$ then (4.26) and (4.27) hold and $T \leq H_{a'}$ for all $a' \in \mathcal{A}$.

Lemma 4.19. *Suppose that for all stopping times S with $S \leq T$ and $\mathbb{E}|B_S| < \infty$ we have*

$$\mathbb{E}(B_T|\mathcal{F}_S) \leq B_S \quad a.s.. \tag{4.28}$$

Then T is minimal.

We refer the reader back to Lemma 3.13, the proof of which is still valid in the more general case.

Of course we may replace the '$\leq$' in (4.28) with '$\geq$' or '$=$' without altering the conclusion.

Lemma 4.20. *If $T \leq H_{\mathcal{A}} = \inf\{t \geq 0 : B_t \in \mathcal{A}\}$ is a stopping time of the Brownian motion $(B_t)_{t\geq 0}$ where $B_0 \sim \mu_0$ and $B_T \sim \mu$, and*

$$\mathbb{E}(B_T|\mathcal{F}_S) \leq B_S : \text{ on } \{B_0 \geq a_-\} \tag{4.29}$$

$$\mathbb{E}(B_T|\mathcal{F}_S) \geq B_S : \text{ on } \{B_0 \leq a_+\}, \tag{4.30}$$

where $a_- = \inf \mathcal{A}$ and $a_+ = \sup \mathcal{A}$, then T is minimal.

Proof. Choose $a \in \mathcal{A}$. By assumption $T \leq H_a$ and by Lemma 4.19 T is minimal for $\check{\mu}^a$ on $\{B_0 \leq a\}$ and for $\hat{\mu}^a$ on $\{B_0 \geq a\}$. It must then be minimal for μ. $\square$

These results show the equivalence of minimality and the conditions in (4.29), (4.30). The following theorem states this together with some extra equivalent conditions. It should be thought of as the extension of Theorem 3.7 to the setting with a general starting measure.

Theorem 4.21. *Let B be a Brownian motion such that $B_0 \sim \mu_0$ and T a stopping time such that $B_T \sim \mu$, where μ_0, μ are integrable. Let $\mathcal{A}$ be the set defined in* (4.17) *and* $a_+ = \sup\{x \in [-\infty, \infty] : x \in \mathcal{A}\}$, $a_- = \inf\{x \in [-\infty, \infty] : x \in \mathcal{A}\}$. *Then the following are equivalent:*

(i) T *is minimal;*

(ii) $T \leq H_{\mathcal{A}}$ *and for all stopping times* $R \leq S \leq T$

$$\begin{aligned} \mathbb{E}(B_S|\mathcal{F}_R) &\leq B_R \text{ on } \{B_0 \geq a_-\} \\ \mathbb{E}(B_S|\mathcal{F}_R) &\geq B_R \text{ on } \{B_0 \leq a_+\}; \end{aligned}$$

(iii) $T \leq H_{\mathcal{A}}$ *and for all stopping times* $S \leq T$

$$\begin{aligned} \mathbb{E}(B_T|\mathcal{F}_S) &\leq B_S \text{ on } \{B_0 \geq a_-\} \\ \mathbb{E}(B_T|\mathcal{F}_S) &\geq B_S \text{ on } \{B_0 \leq a_+\}; \end{aligned}$$

(iv) $T \leq H_{\mathcal{A}}$ *and for all* $\gamma > 0$

$$\begin{aligned} \mathbb{E}(B_T; T > H_{-\gamma}, B_0 \geq a_-) &\leq -\gamma\mathbb{P}(T > H_{-\gamma}, B_0 \geq a_-) \\ \mathbb{E}(B_T; T > H_{\gamma}, B_0 \leq a_+) &\geq \gamma\mathbb{P}(T > H_{\gamma}, B_0 \leq a_+); \end{aligned}$$

(v) $T \leq H_{\mathcal{A}}$ *and as* $\gamma \to \infty$

$$\begin{aligned} \gamma\mathbb{P}(T > H_{-\gamma}, B_0 \geq a_-) &\to 0 \\ \gamma\mathbb{P}(T > H_{\gamma}, B_0 \leq a_+) &\to 0. \end{aligned}$$

We begin by proving the following result:

Proposition 4.22. *If (v) holds and* $S \leq T$ *then* $\mathbb{E}|B_S| < \infty$.

Proof. We show that $\mathbb{E}(|B_S|; B_0 \geq a_-) < \infty$. Since $B_{t\wedge H_{-k}}$ is a supermartingale on $\{B_0 \geq -k\}$,

$$\begin{aligned} &\mathbb{E}(B_{T\wedge H_{-k}}; B_S < 0, S < H_{-k}, B_0 \geq a_- \wedge (-k)) \\ &\qquad \leq \mathbb{E}(B_{S\wedge H_{-k}}; B_S < 0, S < H_{-k}, B_0 \geq a_- \wedge (-k)). \end{aligned}$$

The term on the left hand side is equal to:

$$\begin{aligned} &\mathbb{E}(B_T; B_S < 0, T < H_{-k}, B_0 \geq a_- \wedge (-k)) \\ &\qquad - k\mathbb{P}(B_S < 0, S \leq H_{-k} < T, B_0 \geq a_- \wedge (-k)). \end{aligned}$$

The first term converges (by dominated convergence) to $\mathbb{E}(B_T; B_S < 0, B_0 \geq a_-)$ and the second term vanishes by the assumption. By monotone convergence

$$
\begin{aligned}
\mathbb{E}(B_S; B_S < 0, B_0 \geq a_-) &= \lim_k \mathbb{E}(B_S; B_S < 0, S < H_{-k}, B_0 \geq a_- \wedge (-k)) \\
&\geq \lim_k \mathbb{E}(B_T; B_S < 0, S < H_{-k}, B_0 \geq a_- \wedge (-k)) \\
&\geq \mathbb{E}(B_T; B_S < 0, B_0 \geq a_-) \geq -\mathbb{E}(B_T^-) > -\infty.
\end{aligned}
$$

Also

$$
\begin{aligned}
\mathbb{E}(B_0; B_0 \geq a_- \wedge (-k)) &\geq \mathbb{E}(B_{S \wedge H_{-k}}; B_0 \geq a_- \wedge (-k)) \\
&= \mathbb{E}(B_S; B_0 \geq a_- \wedge (-k), S < H_{-k}) \\
&\quad - k\mathbb{P}(H_{-k} \leq S, B_0 \geq a_- \wedge (-k)),
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{E}(B_S; B_0 \geq a_- \wedge (-k), S < H_{-k}) &= \mathbb{E}(B_S^+; B_0 \geq a_- \wedge (-k), S < H_{-k}) \\
&\quad - \mathbb{E}(B_S^-; B_0 \geq a_- \wedge (-k), S < H_{-k}),
\end{aligned}
$$

so

$$
\begin{aligned}
\mathbb{E}(B_S^+; B_0 \geq a_- \wedge (-k), S < H_{-k}) &\leq \mathbb{E}(B_0; B_0 \geq a_- \wedge (-k)) \\
&\quad + \mathbb{E}(B_S^-; B_0 \geq a_- \wedge (-k), S < H_{-k}) \\
&\quad + k\mathbb{P}(H_{-k} \leq S, B_0 \geq a_- \wedge (-k)).
\end{aligned}
$$

By monotone and dominated convergence, in the limit we have

$$
\begin{aligned}
\mathbb{E}(B_S^+; B_0 \geq a_-) &\leq \mathbb{E}(B_0; B_0 \geq a_-) + \mathbb{E}(B_S^-; B_0 \geq a_-) \\
&< \infty.
\end{aligned}
$$

So $\mathbb{E}(|B_S|; B_0 \geq a_-) < \infty$. Similarly $\mathbb{E}(|B_S|; B_0 \leq a_+) < \infty$, and together these imply $\mathbb{E}(B_S) < \infty$. □

Proof of Theorem 4.21. Clearly *(ii)* $\implies$ *(iii)* $\implies$ *(iv)* $\implies$ *(v)* (the final implication following from dominated convergence). We also know *(i)* $\iff$ *(iii)*. We show *(v)* $\implies$ *(ii)*.

Suppose $A \in \mathcal{F}_R$, $A \subseteq \{B_0 \geq a_-\}$ and set $A_k = A \cap \{R < H_{-k}\} \cap \{B_0 \geq -k\}$. Then

$$\mathbb{E}(B_{S\wedge H_{-k}}; A_k) \leq \mathbb{E}(B_{R\wedge H_{-k}}; A_k).$$

By Proposition 4.22 we may apply dominated convergence to deduce that in the limit as $k \to \infty$ the right-hand side converges to $\mathbb{E}(B_R; A)$. Also

$$\begin{aligned} \mathbb{E}(B_{S\wedge H_{-k}}; A_k) =& \mathbb{E}(B_S; A \cap \{B_0 \geq -k\} \cap \{S \leq H_{-k}\}) \\ &+ k\mathbb{P}(A, R < H_{-k} < S, B_0 \geq -k), \end{aligned}$$

where the second term converges to zero by assumption and the first converges to $\mathbb{E}(B_S; A)$ by dominated convergence. $\square$

4.6 Minimality of the Limit

We will want to show that stopping times constructed using the techniques of Sections 4.2 and 4.3 are indeed minimal when (4.16) is satisfied. To deduce that a stopping time T constructed using the balayage techniques is minimal, we approximate T by the sequence of stopping times T_n given in the construction (so T_1 is the exit time from the first interval we construct, and so on). Then it is clear that the stopping times T_n satisfy the conditions of Lemma 4.20, since they are either the first exit time from a bounded interval, or the first time to leave $(-\infty, \alpha]$ for some α. Our aim is then to deduce that the limit is minimal. We shall do this by extending Proposition 3.18 to the case of a general starting measure.

Proposition 4.23. *Suppose that T_n embeds μ_n, μ_n converges weakly to μ and $\mathbb{P}(|T_n - T| > \varepsilon) \to 0$ for all $\varepsilon > 0$. Then T embeds μ.*

If also $l_n \to l_\infty < \infty$ where $l_n = \int |x|\mu_n(dx)$ and $l_\infty = \int |x|\mu(dx)$, and T_n is minimal for μ_n, then T is minimal for μ.

Remark 4.24. Since $\mu_n \implies \mu$, on some probability space we are able to find random variables X_n and X with laws μ_n and μ such that $X_n \to X$ a.s.. By Scheffé's Lemma therefore

$$\mathbb{E}|X_n - X| \to 0 \text{ if and only if } \mathbb{E}|X_n| \to \mathbb{E}|X|,$$

the second statement being equivalent to $l_n \to l_\infty$ in the statement of Proposition 4.23

Before we prove this result, we will show a useful result on the distribution of the maximum — an extension of Theorem 3.20. This will be used in the proof of the above result, and also be important for the work in the next section, when we will show that the inequality in (4.31) can be attained by a class of stopping times created by balayage techniques.

Lemma 4.25. *Let T be a minimal embedding of μ in a Brownian motion started with distribution μ_0. Then for all $x \in \mathbb{R}$*

$$\mathbb{P}(\overline{B}_T \geq x) \leq \inf_{\lambda < x} \frac{1}{2}\left[1 + \frac{u_{\mu_0}(x) - c(\lambda)}{x - \lambda}\right]. \tag{4.31}$$

Proof. We note the following inequality, which (by considering on a case by case basis) holds for all paths and all pairs $\lambda < x$:

$$\mathbf{1}_{\{\overline{B}_T \geq x\}} \leq \frac{1}{x-\lambda}\left[B_{T\wedge H_x} + \frac{|B_T - \lambda| - (B_T + \lambda)}{2} - \frac{|B_0 - x| + (B_0 - x)}{2}\right]. \tag{4.32}$$

In particular, on $\{\overline{B}_T < x\}$, when therefore $\{B_0 < x\}$:

$$0 \leq \frac{1}{x-\lambda}\left[B_T + \begin{Bmatrix} -\lambda & : B_T > \lambda \\ -B_T & : B_T \leq \lambda \end{Bmatrix}\right]. \tag{4.33}$$

While on $\{\overline{B}_T \geq x\}$,

$$\begin{aligned} 1 &\leq \frac{1}{x-\lambda}\left[B_{T\wedge H_x} + \begin{Bmatrix} -\lambda & : B_T > \lambda \\ -B_T & : B_T \leq \lambda \end{Bmatrix} - \begin{Bmatrix} B_0 - x & : B_0 > x \\ 0 & : B_0 \leq x \end{Bmatrix}\right] \\ &\leq \frac{1}{x-\lambda}\left[x + \begin{Bmatrix} -\lambda & : B_T > \lambda \\ -B_T & : B_T \leq \lambda \end{Bmatrix}\right]. \end{aligned} \tag{4.34}$$

So we may take expectations in (4.32) to get

$$\mathbb{P}(\overline{B}_T \geq x) \leq \frac{1}{2}\left[1 + \frac{2\mathbb{E}(B_{T\wedge H_x}) + (u_{\mu_0}(x) - u_\mu(\lambda)) - (\mathbb{E}(B_T) + \mathbb{E}(B_0))}{(x-\lambda)}\right]. \tag{4.35}$$

We can deduce (4.31) provided we can show

$$C \geq 2\mathbb{E}(B_{T\wedge H_x}) - (\mathbb{E}(B_T) + \mathbb{E}(B_0)) \tag{4.36}$$

since (4.35) holds for all $\lambda < x$.

We now consider $a \in \mathcal{A}$ possibly taking the values $\pm\infty$. Since $u_\mu(a) - u_{\mu_0}(a) = C$ for

$a \in \mathcal{A}$, we can deduce

$$C = 2\mathbb{E}(B_T; B_T \geq a) + 2\mathbb{E}(B_0; B_0 < a) - \mathbb{E}(B_T) - \mathbb{E}(B_0)$$

where we note that $\{B_T < a\} = \{B_0 < a\}$. Theorem 4.21 tells us that

$$\begin{aligned} \mathbb{E}(B_{T\wedge H_x}; B_0 < a) &\leq \mathbb{E}(B_T; B_0 < a) && (4.37)\\ \mathbb{E}(B_{T\wedge H_x}; B_0 \geq a) &\leq \mathbb{E}(B_0; B_0 \geq a) && (4.38)\end{aligned}$$

and (4.36) holds. □

We also have the following result:

Proposition 4.26. *Suppose μ and $\{\mu_n\}_{n\geq 1}$ are all integrable distributions such that $\mu_n \implies \mu$ and $l_n = \int |y|\,\mu_n(dy) \to \int |y|\,\mu(dy) = l_\infty$. Then u_{μ_n} converges uniformly to u_μ.*

Proof. Fix $\varepsilon > 0$. By (4.19), using the fact that $\mu - \hat{\mu} = \check{\mu}$ we may write

$$u_\mu(x) = \int_{-\infty}^{\infty} (x-y)\,\mu(dy) + 2\int_{-\infty}^{x} (y-x)\,\mu(dy) = x - \int_{-\infty}^{\infty} y\,\mu(dy) + 2\int_{-\infty}^{x} (y-x)\,\mu(dy),$$

and similarly for u_{μ_n}, hence

$$u_{\mu_n}(x) - u_\mu(x) = (m_\infty - m_n) + 2\int_{-\infty}^{x} (y-x)\,(\mu_n - \mu)(dy), \tag{4.39}$$

where we write m_n, m_∞ for the means of μ_n and μ respectively; $m_n \to m$ as a consequence of Remark 4.24. Since μ is integrable, as $x \downarrow -\infty$,

$$\int_{-\infty}^{x} (x-y)\,\mu(dy) \downarrow 0.$$

By (4.39) and Lemma 4.3 (which implies u_{μ_n} converges to u_μ pointwise, the C in (4.5) being 0 since $l_n \to l_\infty$), for all $x \in \mathbb{R}$

$$\int_{-\infty}^{x} (x-y)\mu_n(dy) \to \int_{-\infty}^{x} (x-y)\mu(dy)$$

as $n \to \infty$. Finally we note that both sides of the above are increasing in x.

Consider

$$|u_{\mu_n}(x) - u_\mu(x)| \leq |m_\infty - m_n| + 2\int_{-\infty}^{x} (x-y)\,\mu_n(dy) + 2\int_{-\infty}^{x} (x-y)\,\mu(dy).$$

We may choose x_0 sufficiently small that $\int_{-\infty}^{x_0} (x_0 - y)\,\mu(dy) < \varepsilon$, and therefore such that

$$\int_{-\infty}^{x} (x-y)\,\mu(dy) \leq \int_{-\infty}^{x_0} (x_0-y)\,\mu(dy) < \varepsilon$$

for all $x \leq x_0$. By the above and Remark 4.24 we may now choose $n_0(\varepsilon)$ such that for all $n \geq n_0(\varepsilon)$

$$|m_\infty - m_n| < \varepsilon \text{ and } \left| \int_{-\infty}^{x_0} (x_0-y)\,\mu_n(dy) - \int_{-\infty}^{x_0} (x_0-y)\,\mu(dy) \right| < \varepsilon.$$

Then for all $x \leq x_0$ and for all $n \geq n_0(\varepsilon)$,

$$|u_{\mu_n}(x) - u_\mu(x)| \leq \varepsilon + 2 \times 2\varepsilon + 2\varepsilon = 7\varepsilon.$$

Similarly we can find $x_1, n_1(\varepsilon)$ such that $|u_{\mu_n}(x) - u_\mu(x)| \leq 7\varepsilon$ for all $x \geq x_1$ and all $n \geq n_1(\varepsilon)$. Finally u_{μ_n}, u_μ are both Lipschitz and pointwise $u_{\mu_n}(x) \to u_\mu(x)$ and we must have uniform convergence on any bounded interval, and in particular on $[x_0, x_1]$. □

Proof of Proposition 4.23. Suppose first that there exists $a \in \mathcal{A} \cap \mathbb{R}$. We show that $T \leq H_a$ for all such a. As usual, we write μ_0 for the starting measure, and $c(x) = u_\mu(x) - C$. We define C_n to be the smallest value such that $u_{\mu_0}(x) \geq u_{\mu_n}(x) - C_n$ and the functions $c_n(x) = u_{\mu_n}(x) - C_n$. Note that $l_n = u_{\mu_n}(0)$, so $\lim_{n\to\infty} u_{\mu_n}(0)$ exists. Then (by Lemma 4.3*(i)* or equivalently (Chacon, 1977)[Lemma 2.5]) weak convergence implies

$$\lim_{n\to\infty} u_{\mu_n}(x) = u_\mu(x) - K$$

for all $x \in \mathbb{R}$ and (here) $K = 0$ since $u_{\mu_n}(0) \to u_\mu(0)$.

By Lemma 4.25 for $x \in \mathbb{R}$ and $\lambda < x$

$$\mathbb{P}(\overline{B}_{T_n} \geq x) \leq \frac{1}{2}\left[1 + \frac{u_{\mu_0}(x) - u_{\mu_n}(\lambda) + C}{x - \lambda} + \frac{C_n - C}{x - \lambda}\right],$$

and we take the limit as $n \to \infty$, using Proposition 4.26 (so that $C_n \to C$) and noting that $\mathbb{P}(\overline{B}_{T_n} \geq x) \to \mathbb{P}(\overline{B}_T \geq x)$, to get

$$\mathbb{P}(\overline{B}_T \geq x) \leq \frac{1}{2}\left[1 + \frac{u_{\mu_0}(x) - c(\lambda)}{x - \lambda}\right].$$

Suppose now $x = a$. Since the above holds for all $\lambda < a$, we may take the limit of the right hand side as $\lambda \uparrow a$, in which case $u_{\mu_0}(a) = c(a)$, and by Remark 4.2

$$\begin{aligned}\mathbb{P}(\overline{B}_T \geq a) &\leq \frac{1}{2}\left[1 + c'_-(a)\right] \\ &\leq \frac{1}{2}\left[1 + (1 - 2\mu((-\infty, a)))\right] \\ &\leq \mu([a, \infty)).\end{aligned}$$

By considering $-B_t$ we may deduce that $\mathbb{P}(\underline{B}_T \leq a) \leq \mu((-\infty, a])$. Hence $\mathbb{P}(T \leq H_a) = 1$, and we deduce that T is minimal.

It only remains to show (by Lemma 4.19) that if $\infty \in \mathcal{A}$ then

$$\mathbb{E}(B_T | \mathcal{F}_S) \geq B_S$$

for all stopping times $S \leq T$. The case where $-\infty \in \mathcal{A}$ follows from $B_t \mapsto -B_t$. In particular, for $S \leq T$ and $A \in \mathcal{F}_S$ we need to show

$$\mathbb{E}(B_T; A) \geq \mathbb{E}(B_S; A). \tag{4.40}$$

In fact we need only show the above for sets $A \subseteq \{S < T\}$ since it clearly holds on $\{S = T\}$. So we can define $A_n = A \cap \{S < T_n\}$ and therefore $\mathbb{P}(A \setminus A_n) \to 0$ as $n \to \infty$. Also $A_n \in \mathcal{F}_{S \wedge T_n}$. By Theorem 4.21 and the fact that the T_n are minimal

$$\begin{aligned}\mathbb{E}(B_{S \wedge T_n}; A_n) &\leq \mathbb{E}(B_{T_n}; A_n \cap \{B_0 \leq a_+^n\}) + \mathbb{E}(B_{S \wedge T_n}; B_0 > a_+^n) \\ &\quad - \mathbb{E}(B_{S \wedge T_n}; A_n^C \cap \{B_0 > a_+^n\}) \\ &\leq \mathbb{E}(B_{T_n}; A_n \cap \{B_0 \leq a_+^n\})\mathbb{E}(B_0; B_0 > a_+^n) \\ &\quad - \mathbb{E}(B_{T_n}; A_n^C \cap \{B_0 > a_+^n\}) \\ &\leq \mathbb{E}(B_{T_n}; A_n) - \mathbb{E}(B_{T_n}; \{B_0 > a_+^n\}) + \mathbb{E}(B_0; B_0 > a_+^n)\end{aligned}$$

where a_+^n is the supremum of the set $\mathcal{A}_n$ (that is the corresponding set to $\mathcal{A}$ for the measures μ_0, μ_n). This is not necessarily infinite.

So it is sufficient for us to show that

$$\begin{aligned}\lim_n \mathbb{E}(B_{T_n}; A_n) &= \mathbb{E}(B_T; A); && (4.41)\\ \lim_n \mathbb{E}(B_S; A_n) &= \mathbb{E}(B_S; A), && (4.42)\end{aligned}$$

and

$$\lim_n |\mathbb{E}(B_0; B_0 > a_+^n) - \mathbb{E}(B_{T_n}; B_0 > a_+^n)| = 0. \tag{4.43}$$

For (4.41) we may use a proof identical to that used in Proposition 3.18 to show (3.23). We want to apply Lemma 4.19 so we can assume that $\mathbb{E}|B_S| < \infty$, and (4.42) follows by dominated convergence.

Finally we consider (4.43). Let $\theta_n = \mu_0((-\infty, a_+^n])$. Since $a_+^n \in \mathcal{A}_n$ we have

$$\begin{aligned}\mathbb{E}(B_0; B_0 > a_+^n) &- \mathbb{E}(B_{T_n}; B_0 > a_+^n)\\ &= \int y\,\hat{\mu}_0^{a_+^n,\theta_n}(dy) - \int y\,\hat{\mu}_n^{a_+^n,\theta_n}(dy)\\ &= \int (y - a_+^n)\,\hat{\mu}_0^{a_+^n,\theta_n}(dy) - \int (y - a_+^n)\,\hat{\mu}^{a_+^n,\theta_n}(dy)\\ &= \frac{1}{2}\left[\int y\,(\mu_0 - \mu_n)(dy) + u_{\mu_n}(a_+^n) - u_{\mu_0}(a_+^n)\right]\\ &= \frac{1}{2}\left[\int y\,(\mu_0 - \mu)(dy) - C_n\right],\end{aligned}$$

where we have used the fact that (for a general measure ν)

$$\int (y - x)\,\hat{\nu}^x(dy) = \frac{1}{2}\left[\int y\,\nu(dy) - u_\nu(x) - x\right].$$

As $n \to \infty$, since $\infty \in \mathcal{A}$,

$$\int y\,(\mu_0 - \mu_n)(dy) \to \int y\,(\mu_0 - \mu)(dy) = C.$$

So we need only show that $C_n \to C$, which follows from the uniform convergence of u_{μ_n} to u_μ (Proposition 4.26). $\square$

4.7 Tangents and Azema-Yor Type Embeddings

One of the motivations for this chapter is to discuss generalisations of the Azema-Yor family of embeddings (see Azéma and Yor (1979a); Jacka (1988) and Chapter 3) to the integrable starting/target measures we have discussed already.

The aim is therefore to find the embedding which maximises the law of the maximum, $\sup_{0 \le t \le T} B_t$ (or in the more general case $\sup_{0 \le t \le T} f(B_t)$). If we look for the maximum within the class of all embeddings there is no natural maximum embedding. For this reason we consider the class of minimal embeddings. Lemma 4.25 establishes that there is some natural limit when we consider this restriction. In fact the extended Azema-Yor embedding will attain the limit in (4.31).

The idea is to use the machinery from the previous sections to show the embeddings exist as limits of the Chacon-Walsh type embeddings of Section 4.3. It is then possible to show that the embeddings are minimal and that they attain equality in (4.31).

Theorem 4.27. *If T is a stopping time as described in Lemma 4.9, where C as described in the lemma is*

$$C = \inf_x \{u_\mu(x) - u_{\mu_0}(x)\}, \tag{4.44}$$

then T is minimal.

Proof. Lemma 4.9 suggests a sequence T_n of stopping times for which T is the limit. We note that we can modify the definition of T_n so that T_n' is specified by the functions $f_1, f_2, \ldots, f_n, f^{-1}, f^{+1}$ without altering their limit (as a consequence of (4.11)), where f^{-1} is the tangent to g with gradient -1 and f^{+1} is the tangent to g with gradient 1. It is easy to see that this ensures that $\mathbb{E}(B_{T_n'}) = \mathbb{E}(B_T)$ (by (4.4)), and also that $u_{\mu_n}(0) \to u_\mu(0)$ and $n \to \infty$. Consequently the stopping times T_n' and T satisfy the conditions of Proposition 4.23, where it is clear that the T_n' are all minimal, since each step clearly satisfies the conditions of Theorem 4.21 as a consequence of (4.44). So T is minimal. □

Define the function

$$\Phi(x) = \operatorname*{argmin}_{\lambda < x} \left\{ \frac{u_{\mu_0}(x) - c(\lambda)}{x - \lambda} \right\}. \tag{4.45}$$

In the cases described by Azéma and Yor (1979*a*), this is the barycentre function. It can also be seen to agree with the function appearing in the generalisation of the Azema-Yor stopping time to non-centred means which appears in (3.26). A similar function is used in Hobson (1998*a*) who examines the case where starting and target means are centred and satisfy (4.10). $\Phi(\cdot)$ can be thought of graphically as the point (below x) at which there exists a tangent to $c(\cdot)$ meeting the function $u_{\mu_0}(\cdot)$ at x.

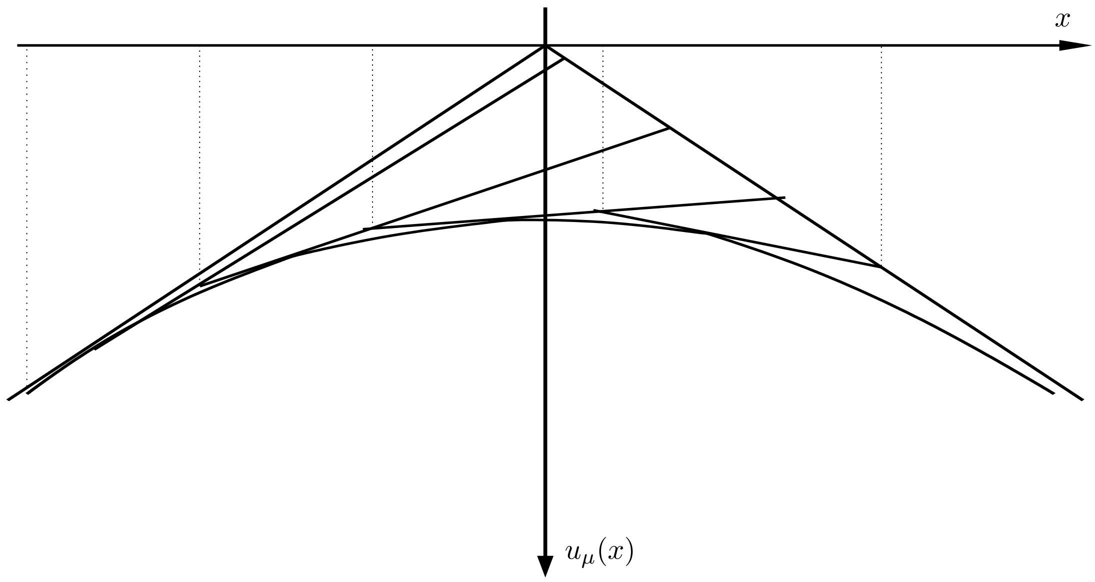

Figure 4-3: Approximating the Azema-Yor stopping time: we take tangents to the potential from left to right. In the limit the tangents become closer. The dotted lines highlight the points at which the approximated stopping time will stop the process.

Lemma 4.28. *The Azema-Yor stopping time*

$$T = \inf\{t \geq 0 : B_t \leq \Phi(\overline{B}_t)\} \tag{4.46}$$

is minimal and attains equality in (4.31)*.*

We prove this lemma using an extension of an idea first suggested in Meilijson (1983). We approximate T by taking tangents to c, starting with gradient -1, and increasing to $+1$. As the number of tangents we take increases, the stopping time converges to T. The general approximation sequence can be seen in Figure 4-3.

Proof. We apply Lemma 4.9 for each n to the functions $f_1^n, f_2^n, \ldots, f_{m(n)}^n$, which are chosen as tangents to $c(\cdot)$ with increasing gradients, so that f_1^n has gradient -1, f_m^n has gradient 1, and so that the difference in the gradients of consequential tangents is less than $\frac{1}{n}$. We also choose the tangents in such a way that the points at which successive tangents intersect each other (which are B_{T_n} stops) are at most $\frac{1}{n}$ apart when they lie within $[-n, n]$ (at least as far as this is possible — if both μ_0 and μ have an interval containing no mass, it might not be possible to manage this, but this case will not be important). This defines a (minimal) stopping time T_n such that (by (4.4)) $\mathbb{E}(B_{T_n}) = \int x\,\mu(dx)$. Also, by considering $\mu_n = \mathcal{L}(B_{T_n})$, $|\mu_n((-\infty, x)) - \mu((-\infty, x))| \leq \frac{1}{n}$ for all

$x \in \mathbb{R}$. So $\mu_n \implies \mu$. The choice of T_n also ensures that $\mathbb{P}(|T - T_n| > \varepsilon) \to 0$ for all $\varepsilon > 0$. Consequently T is minimal.

To deduce that T attains equality in (4.31) we note that $\Phi(x)$ is the optimal choice for λ in (4.31), and by the definition of $\Phi(x)$,

$$\begin{aligned} \{\overline{B}_T < x\} &\subseteq \{B_T \leq \Phi(x)\} \\ \{\overline{B}_T \geq x\} &\subseteq \{B_T \geq \Phi(x)\}. \end{aligned}$$

This means we attain equality in (4.33) and (4.34), and so only need show that we have equality in (4.37) and (4.38) for equality in (4.31) to hold. But for x given, we may calculate the potential of $\mu' = \mathcal{L}(B_{T \wedge \bar{H}_x})$ — where $\bar{H}_x = \inf\{t \geq 0 : B \geq x\}$ — as:

$$u_{\mu'}(y) = \begin{cases} u_\mu(y) & : y \leq \Phi(x); \\ u_\mu(\Phi(x)) + \frac{y-\Phi(x)}{x-\Phi(x)}(u_{\mu_0}(x) - u_\mu(\Phi(x))) & : \Phi(x) \leq y \leq x; \\ u_{\mu_0}(y) & : y \geq x. \end{cases}$$

It then follows from Theorem 4.21 and (4.4) that equality holds. $\square$

4.8 The Vallois Construction

We conclude with a second example demonstrating the advantages of the Chacon-Walsh construction, and its power when used in conjunction with the preceding results. We do much the same as in Section 4.7, in that we construct a sequence of stopping times through balayage for which the desired limit (in the centred case) is the construction first derived in Vallois (1983). In particular it will be comparatively simple to see how the construction extends to both non-centred target distributions and general starting distributions, and it will be a simple consequence that the construction in all these cases is minimal.

Our main emphasis is on showing that the Vallois construction is a special case of the balayage construction introduced earlier in the chapter. In this sense, the comparison should be made with the work of Meilijson (1983), who showed that the Azema-Yor stopping time is a special case of the Chacon-Walsh construction. We do not intend to give a rigorous exposition, but we hope that the discussion here is sufficient to convince the reader that the connection between the Vallois stopping time and the construction we give in this section is valid.

For simplicity we assume that our target distributions have a density with respect to Lebesgue measure, so that c is a twice differentiable function. We also suppose that our starting distribution is the unit mass at 0. We will discuss later the consequences of a general starting distribution.

The Vallois construction can be described as follows: given a centred target distribution μ there exist non-negative, non-increasing functions h, k such that the stopping time

$$T_V = \inf\{t \geq 0 : B_T \notin (-h(L_t), k(L_t))\} \tag{4.47}$$

is an embedding of μ, where L_t is the local time at 0. Vallois (1992) demonstrates also that the embedding maximises the law of the local time among the class of UI embeddings. A corresponding stopping time also exists where the functions h, k are non-decreasing which minimises the law of the local time.

Our aim is to approximate T_V using an appropriate sequence T_m. A key idea in this approximation is that of downcrossings. Specifically, for $\varepsilon > 0$, we define recursively

$$\begin{aligned} R_0^\varepsilon &= 0; \\ S_n^\varepsilon &= \inf\{t > R_n^\varepsilon : B_t = \varepsilon\}, \qquad n \geq 0; \\ R_n^\varepsilon &= \inf\{t > S_{n-1}^\varepsilon : B_t = 0\}, \qquad n \geq 1; \end{aligned}$$

and the number of downcrossings at time t of the interval $[0, \varepsilon]$ is then defined to be:

$$d_\varepsilon(t) = \max\{n : R_n^\varepsilon < t\}.$$

Then the following theorem links the number of downcrossings to the local time:

Theorem 4.29 (Revuz and Yor (1999) Ch. VI, 1.10)**.** *If T is a stopping time of a Brownian motion such that for $p \geq 1$*

$$\mathbb{E}T^{p/2} < \infty,$$

then

$$\lim_{\varepsilon \to 0} \mathbb{E}\left[\sup_{t \leq T}\left|\varepsilon d_\varepsilon(t) - \frac{1}{2}L_t\right|^p\right] = 0.$$

Graphically our stopping times can be described in the Chacon-Walsh sense as follows (see Figure 4-4). Let $\varepsilon(m) > 0$ be a decreasing sequence, so that $\varepsilon(m) \downarrow 0$ as $m \to \infty$. For each ε we construct tangents to c so that the first tangent (tangential to c at some point less than 0) passes through $(\varepsilon, -\varepsilon)$, the second tangent (tangential to c at some

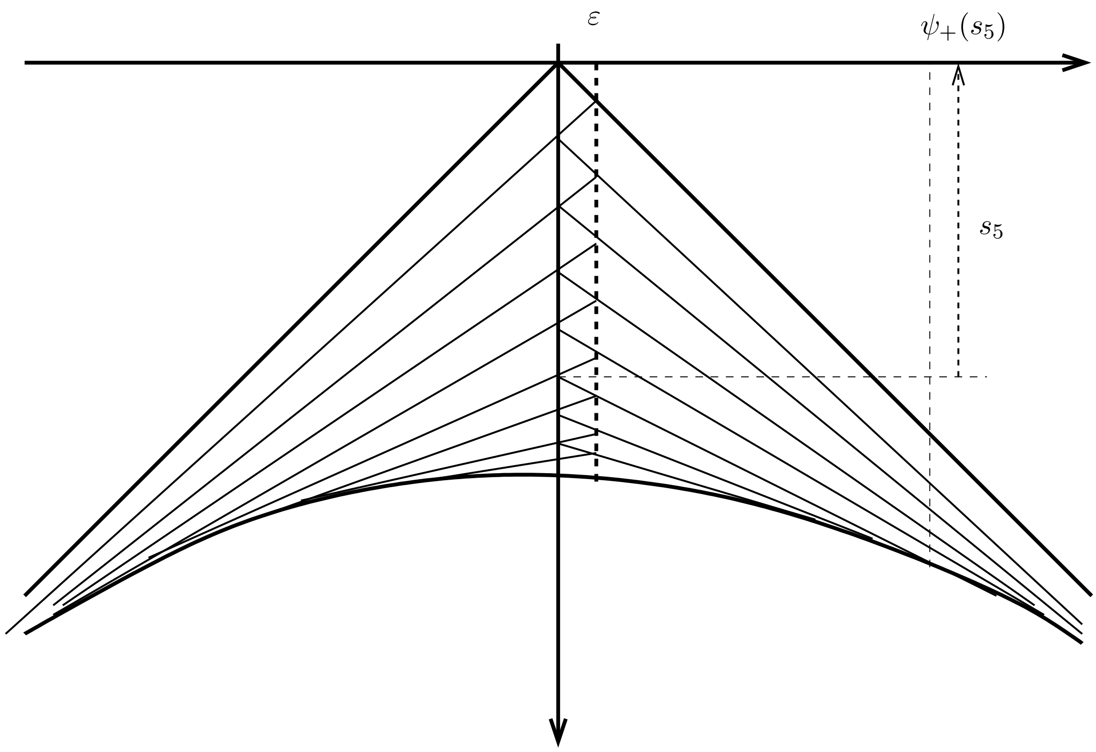

Figure 4-4: The Chacon-Walsh type picture for an approximation to the Vallois stopping time. In the limit, we allow $\varepsilon \to 0$.

positive point) passes through the intersection of the previous tangent and the line $x = 0$; the third tangent is now chosen to intersect the second tangent and the line $x = \varepsilon$ at the same point. This procedure is repeated as far as possible. In terms of when we stop the process, there is the following interpretation: starting from 0 we run the process until it hits ε or some lower level (depending on the current number of downcrossings); from ε the process then runs until it either hits some upper level (again depending on the number of downcrossings already made) or it returns to 0, having made one more downcrossing.

From the picture we can see the following quantities will be important: for $s \geq 0$ define

$$\begin{aligned} \psi_-(s) &= \sup\{x \leq 0 : c(x) - xc'(x) = -s\}; \\ \psi_+(s) &= \inf\{x \geq 0 : c(x) - xc'(x) = -s\}. \end{aligned}$$

We now make the construction of T_m explicit: we write ε for $\varepsilon(m)$, and define recursively

$s_0, s_1, \ldots$ to be the (unique) solutions of the equations

$$\begin{aligned} s_0 &= 0 \\ s_1 &= \varepsilon + \varepsilon \frac{c(\psi_-(s_1)) + s_1}{\psi_-(s_1)} \\ &\vdots \\ s_n &= s_{n-1} + \varepsilon \left[\frac{c(\psi_-(s_n)) + s_n}{\psi_-(s_n)} - \frac{c(\psi_+(s_{n-1})) + s_{n-1}}{\psi_+(s_{n-1})} \right]. \end{aligned} \tag{4.48}$$

We note also that

$$c'(\psi_\pm(s)) = \frac{s + c(\psi_\pm(s))}{\psi_\pm(s)} \tag{4.49}$$

and $c'(\psi_+(0)) = -1$ so that we may also write

$$s_n = \varepsilon \sum_{k=1}^{n} \left[c'(\psi_-(s_k)) - c'(\psi_+(s_{k-1})) \right].$$

We repeat this procedure as far as possible (a finite number of steps), and we stop the process once this is no longer possible. The stopping time T_m can then be defined as

$$T_m = \inf\{t \geq 0 : B_t \notin (\psi_-^m(s_{d_\varepsilon(t)}), \psi_+^m(s_{d_\varepsilon(t)}))\}$$

where $\psi_+^m(s_{d_\varepsilon(t)})$ is actually the x-value of the intersection of the line passing through $(0, -s_{d_\varepsilon(t)})$ and $(\psi_+(s_{d_\varepsilon(t)}), c(\psi_+(s_{d_\varepsilon(t)}))$ and the line passing through $(0, -s_{d_\varepsilon(t)-1})$ and $(\psi_+(s_{d_\varepsilon(t)-1}), c(\psi_+(s_{d_\varepsilon(t)-1}))$. It is clear that as $m \to \infty$, on $[\delta, \infty)$ for any $\delta > 0$ we have uniform convergence $\psi_+^m \to \psi_+$. A similar relation holds for ψ_-^m.

We shall be interested in comparing the limit to T_V, so we need to be more specific about the construction of the functions h, k. Vallois (1983) uses essentially the inverse of the functions ψ_+, ψ_- but can be seen easily to be the same as:

$$T_V = \inf\{t \geq 0 : B_t \notin (\psi_-(2F^{-1}(L_t), \psi_+(2F^{-1}(L_t)))\}$$

where the function F is defined by:

$$\begin{aligned} \lambda(s) &= 1 - \int_0^s \left(\frac{1}{\psi_+(2u)} - \frac{1}{\psi_-(2u)} \right) du; \\ F(s) &= 2 \int_0^s \frac{du}{\lambda(u)}. \end{aligned}$$

Write μ_m for the law of B_{T_m}. Our goal is to apply Proposition 4.23 to the T_m's and show that T_V is indeed their limit, and hence that T_V embeds and is minimal. To apply the result, we need to show the following:

(i) $\mu_m \implies \mu$ as $m \to \infty$;

(ii) $\int |x|\mu_m(dx) \to \int |x|\mu(dx)$ as $m \to \infty$;

(iii) the T_m's are minimal;

(iv) $\mathbb{P}(|T_m - T_V| > \delta) \to 0$ as $m \to \infty$ for all $\delta > 0$.

(i)–(iii) all follow trivially by construction. We need to show $\mathbb{P}(|T_m - T_V| > \delta)$ is sufficiently small for large m. By ignoring an event of small probability we may assume that $s_{d_\varepsilon(T_m)}$ and $F^{-1}(L_{T_V})$ are bounded away from 0 and ∞. Also we may then assume that ψ_+, ψ_- are uniformly continuous, and m is large enough for ψ_+^m, ψ_-^m to approximate ψ_+, ψ_- sufficiently well. Consequently there are essentially two different ways in which we can have $|T_m - T_V| > \delta$:

- $s_{d_\varepsilon(t)}$ and $F^{-1}(L_t)$ are substantially different at some time t;
- $s_{d_\varepsilon(t)}$ and $F^{-1}(L_t)$ are close, but the process stops under T_m or T_V and does not hit the slightly higher level in a short time, possibly even returning to 0 in the intermediate time.

The probability of the second event can be made sufficiently small by ensuring that the points $s_{d_\varepsilon(t)}$ and $F^{-1}(L_t)$ are sufficiently close. So we will be done if we can show that

$$\mathbb{P}\left(\sup_{t \le T_V \vee T_m} \left|s_{d_\varepsilon(t)} - F^{-1}(L_t)\right| > \delta'\right) \to 0$$

as $m \to \infty$. We note however that $\mathbb{E}(T_V \vee T_m)^{1/2} < \infty$, so that by Theroem 4.29

$$\mathbb{P}\left(\sup_{t \le T_V \vee T_m} \left|d_\varepsilon(t) - \frac{1}{2}L_t\right| > \delta''\right) \to 0.$$

So we need to show that $s_{d_\varepsilon(t)} \approx F^{-1}(2\varepsilon d_\varepsilon(t))$, since F^{-1} is uniformly continuous away from 0 and ∞.

Consider $d_{\varepsilon(t)}$:

$$\begin{aligned} d_{\varepsilon(t)} &= \sum_{k=1}^{d_\varepsilon(t)} \frac{s_k - s_{k-1}}{s_k - s_{k-1}} \\ &= \sum_{k=1}^{d_\varepsilon(t)} \frac{s_k - s_{k-1}}{\varepsilon\left[c'(\psi_-(s_k)) - c'(\psi_+(s_{k-1}))\right]}, \end{aligned}$$

where the second line follows from (4.48) and (4.49). Since on letting $m \to \infty$ the s_k become closer, in the limit we would expect the right hand side to approximate $G(s_{d_\varepsilon(t)})$, where we define the function G by

$$G(x) = \int_0^x \frac{du}{c'(\psi_-(u)) - c'(\psi_+(u))}.$$

It therefore just remains to show that $G(2x) = 2F(x)$, however clearly $G(0) = 0 = F(0)$. On differentiating and taking reciprocals we are reduced to showing that

$$\frac{1}{2}\left[c'(\psi_-(2x)) - c'(\psi_+(2x))\right] = 1 - \int_0^x \left(\frac{1}{\psi_+(2x)} - \frac{1}{\psi_-(2x)}\right) dx.$$

Again both sides agree on taking $x = 0$; that they are the same function can be concluded by differentiating and using the relation (4.49).

As already noted, the above construction will produce minimal embeddings for non-centred target distributions (when one of ψ_+ or ψ_- will be infinite for small values), and can be extended to general target distributions in various ways. One of these is depicted in Figure 4-5, the idea being that, as much as possible while keeping the process minimal, we run until we hit zero, with the rest of the mass stopping at the extremes. The mass at zero can then be embedded using the standard Vallois construction, while the mass at the extremes must still be embedded using some other technique — possibly based on the local time at some new level. More generally this technique can be extended so that suitable points $x_1, x_2, \ldots$ are chosen and the process run to hit these points, from which a local-time based procedure can be used. These issues point to the fact that there is no unique natural extension of the Vallois construction to general starting measures; one way of seeing this is to consider an optimality property of the original construction. Vallois (1983) shows that the construction maximises the distribution of the local time at zero; in the general starting distribution example there is mass that cannot be made to reach zero, and so, in terms of maximising the distribution of the local time at zero, the construction we suggest in Figure 4-5 would appear to be optimal but not unique, since any suitable embedding can be chosen for

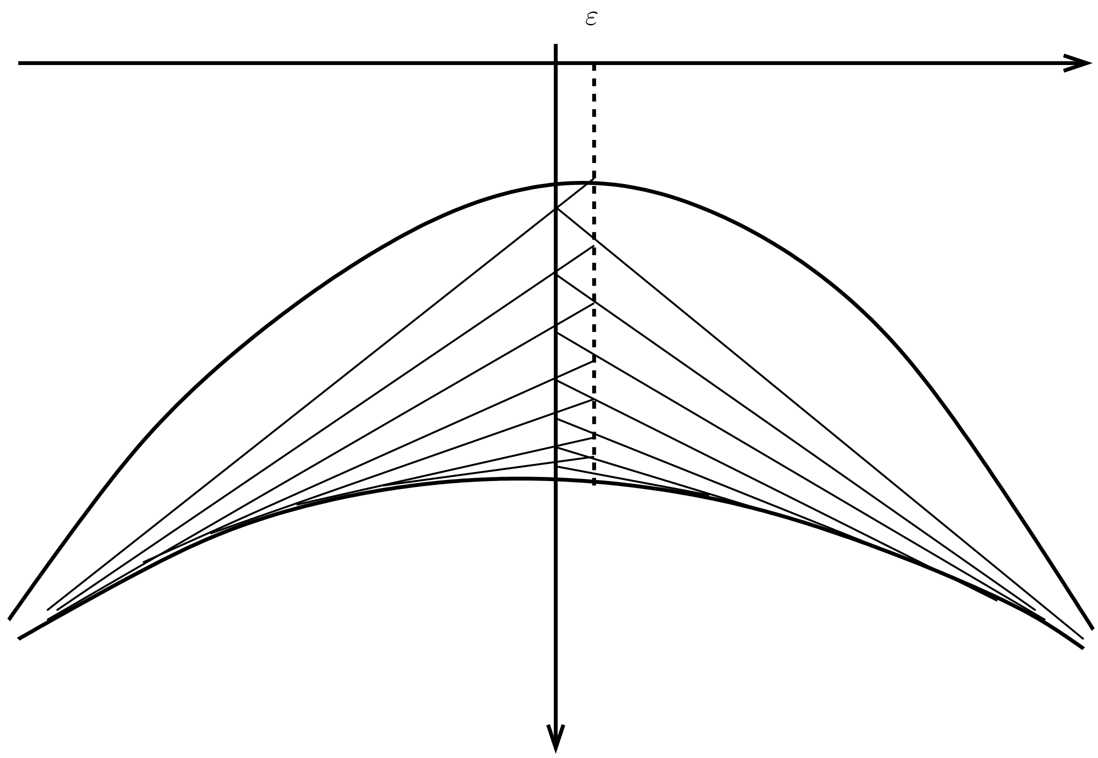

Figure 4-5: The Chacon-Walsh type picture for an approximation to the Vallois stopping time, with a general starting distribution. We note that after the first two steps, there could still be mass at the extremes. This mass will have to be embedded using some suitable procedure — for example a Vallois construction using the local time at a different level.

that part of the process which never hits zero.

Chapter 5

Further Work

In this final chapter we present some further questions which have arisen from the previous work.

In Chapters 3 and 4 we demonstrated that minimality is an important idea when considering which embeddings are suitable, and we were able to give necessary and sufficient conditions for the the process to be minimal when the starting and target distributions are integrable. This leads us to ask what conditions might be necessary and sufficient when the target distribution, and possibly also the starting distribution are not integrable. In this context many of the necessary and sufficient conditions we give in (for example) Theorem 3.7 are no longer necessarily appropriate — many of the conditions are no longer reasonable, for example if the negative tail is not integrable, we cannot always have

$$\gamma\mathbb{P}(T > H_{-\gamma}) \to 0$$

as $\gamma \to \infty$.

In fact even stranger things can happen! If we just consider the case where the negative tail of the distribution is not integrable, but the positive tail is — what we might call the $m = -\infty$ case, we can provide the following example. Suppose we start at zero and have a non-integrable target distribution μ with all its mass placed on $(-\infty, -1)$, we may embed in the following manner: run the process until it hits $+1$, and look at the minimum at this time, the distribution of which may be calculated easily to be

$$\mathbb{P}(\underline{B}_{H_1} \leq -x) = \frac{1}{1 + |x|}$$

for $x \geq 0$. If further we demand that $\mu((-\infty, -x)) > \frac{1}{1+|x|}$ for all $x > 0$ then we can find a function h which is non-decreasing such that $h(B_{H_1})$ has the distribution μ and such that $h(x) < x$. So having hit 1 we run the process until it hits $h(B_{H_1})$, which has the desired distribution, and is such that the process stops at its minimum. This ensures that the stopping time is minimal — any strictly smaller stopping time must have a strictly smaller minimum, but the minimum of any embedding must stochastically dominate the target distribution. This stopping time has the following unexpected property: the stopping time $S = \inf\{t \geq H_1 : B_t = -1\}$ is smaller than T but is not itself minimal; also condition *(ii)* of Theorem 3.7 does not hold — taking $S = H_1$ and any $R \leq S$ contradicts the condition. Condition *(iii)* of the theorem still holds, and could be a necessary and sufficient condition for the stopping time to be minimal when $m = -\infty$; however the example given shows that a proof of this result will be trickier than in the case where the target distribution is integrable, and the case where m is not even defined would seem to be even harder since it is even less clear what the appropriate conditions might be. In the more general case where there is a non-integrable starting distribution, by comparison with Theorem 4.21, we might expect some dependence on the potential, and again here there is a further complication since the potential as we have defined it is only finite for integrable distributions.

The brief discussion of the construction of the Vallois stopping time in Section 4.8 suggests two questions for further research. As mentioned, Vallois (1983) provides two similar constructions of emeddings, both of the form given in (4.47). The one we consider is where the functions h, k are both decreasing, however there is a second embedding in which the functions are chosen to be increasing. While it is possible to see how the functions arise from the potential/Chacon-Walsh picture, in the same way that we do for the decreasing case, there does not appear to be a way of constructing the stopping times by approximating with Chacon-Walsh stopping times. A similar problem can be seen with the stopping times discussed in Perkins (1986) and Chapter 2, where intermediate stages can be interpreted in the potential picture, but it does not appear to be possible to interpret the stopping times as the limit of a balayage construction. One possible explanation for this dichotomy is the fact that both the Perkins embedding and the increasing Vallois case are 'inside-out' embeddings — that is they begin embedding the distribution close to the starting point, and do not embed the extremes until later in the process. This becomes hard to interpret graphically in the Chacon-Walsh picture. The question then becomes: is there a picture in which we can interpret the second Vallois and/or the Perkins embedding as the limit of balayage steps?

The second question that arises from Section 4.8 is how far can we extend the Vallois

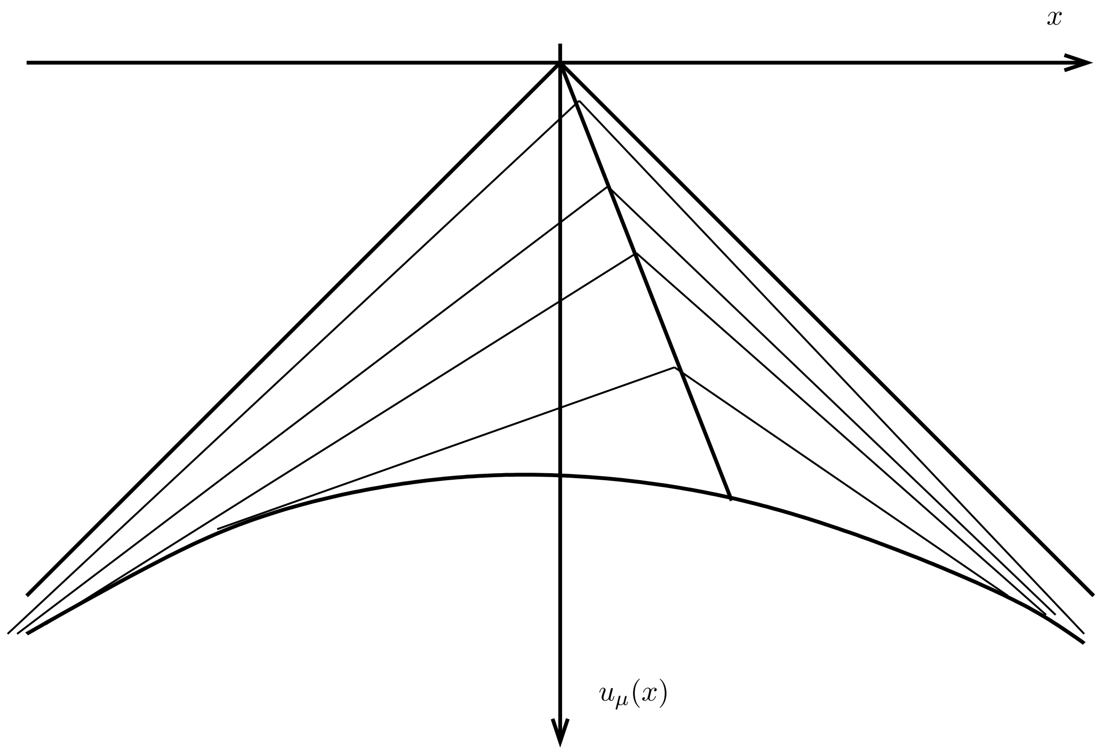

Figure 5-1: The Chacon-Walsh type picture for an extension to the Vallois stopping time. Instead of using the local time at zero we consider approximations along the line $x = ay$, for $-1 < a < 1$. The resulting stopping time appears to be based on the local time of a derived skew Brownian motion. In general it would appear to be possible to replace straight lines with suitable classes of curves.

construction? The pictoral interpretation we give uses the local time along the line $x = 0$. There is no reason why the construction cannot be performed along any line $x = ay$ for $-1 < a < 1$, using a similar procedure to before (see Figure 5-1). This construction appears to have an interpretation in terms of the local time of a skew Brownian motion. A skew Brownian motion can be thought of in excursion terms as a Brownian motion with standard Brownian excursions from zero, but which are negative with probability p and positive with probability $1 - p$. The construction is then similar to the standard Vallois construction, but based on the local time of the derived skew Brownian motion. We note that the construction also extends to the case where $a = 1$ when the skew Brownian motion is really just the excursions of a Brownian motion from its maximum. In this case it would appear that the construction is just the Azema-Yor construction. This idea could be extended even further — rather than just considering straight paths in the potential picture a wider class of paths could possibly be considered.

Another direction in which the ideas surrounding minimality could be considered is in the case of different classes of processes. Two seemingly simple examples of more complicated processes are Brownian motion in higher dimensions, where necessary and sufficient conditions for existence of embeddings are known, but not conditions for minimality; also for example Brownian motion on the circle can be considered. Another seemingly simple case where a variety of issues appear to lie is embedding in a simple symmetric random walk on $\mathbb{Z}$. This case can be easily linked to the Brownian case by considering the walk generated by a given Brownian motion in the obvious way. If we want to construct an embedding we simply construct an embedding for the Brownian case with the target distribution on $\mathbb{Z}$ and carry this over to the random walk example. However in general, for example if we use the Azema-Yor stopping time, in the random walk sense this will involve some independent randomisation. It would seem preferable in the random walk case to have a minimal stopping time not dependent on independent randomisation. If a stopping time is minimal in the Brownian case, it would be minimal in the random walk case, but if we restrict attention only to non-randomised stopping times is it still true that a stopping time that is minimal in the class of non-radomised stopping times is minimal in the class of all stopping times? The answer appears to be no: consider a target distribution with mass $\frac{1}{3}$ at each of $-1, 0, 1$. Allowing randomised stopping times means that the minimal stopping times do not go outside $\{-1, 0, 1\}$, but if we do not allow randomised stopping times this is not possible. Consequently one can ask a variety of questions concerning for example the difference between the class of minimal randomised and non-randomised stopping times of simple random walks.

Appendix A

Some Results From the Introduction

A.1 Some Calculations for Embeddings

A.1.1 The 'Quick and Dirty' Solution

This stopping time is commonly attributed to Doob (see for example Rogers and Williams (2000*a*)[I.7]). We define the supremum and infimum processes of B to be:

$$\begin{aligned}\overline{B}_t &= \sup_{s \leq t} B_t;\\ \underline{B}_t &= \inf_{s \leq t} B_t.\end{aligned}$$

Proposition A.1. *The stopping time T_Q of Example 1.1 has the following properties:*

(i) T_Q is an embedding;

(ii) $\mathbb{E}T_Q = \infty$;

(iii) $\mathbb{E}\overline{B}_{T_Q} \vee \mathbb{E}\underline{B}_{T_Q} = \infty$,

unless μ is the $\mathcal{N}(0,1)$ distribution, when only (i) holds (and $T_Q \equiv 1$).

Proof. $\Phi(B_1) \sim U[0,1]$ so for $x \in \mathbb{R}$

$$\mathbb{P}(F^{-1}(\Phi(B_1)) \leq x) = \mathbb{P}(\Phi(B_1) \leq F(x)) = F(x)$$

and $F^{-1} \circ \Phi(B_1) \sim \mu$. Since B is recurrent $T < \infty$ a.s. and $B_T \sim \mu$.

For the remaining two statements we show that they are true for the stopping time $H_1 = \inf\{t \geq 0 : B_t = 1\}$ of a standard Brownian motion and note that (when μ is not $\mathcal{N}(0,1)$) $F^{-1}(\Phi(B_1)) \neq B_1$ with positive probability. For *(ii)*, it is a well known property of Brownian motion that $\mathbb{E}H_1 \wedge H_{-n} = n$; this stopping time increases almost surely to H_1 so by monotone convergence $\mathbb{E}H_1 = \infty$. Also well known is the fact that $\mathbb{P}(H_{-x} < H_1) = \frac{1}{1+x}$ so that

$$\mathbb{E}(\underline{B}_{H_1}) = \int_0^\infty \frac{1}{1+x}\,dx = \infty,$$

and $\mathbb{E}(\overline{B}_{H_{-1}}) = \infty$. Of course these results hold for H_x for all $x \in \mathbb{R} \setminus \{0\}$ and hence for T_Q. □

A.1.2 Skorokhod's Solution

The stopping time given in Skorokhod (1965) is in fact slightly different to the one we give here (in the choice of ν) — he uses a deterministic relationship between X and Y. The properties of the two embeddings are identical.

Proposition A.2. *Let μ be a centred distribution. The stopping time T_S of* (1.1)*, where ν is defined to be*

$$\nu(A_1 \times A_2) = \int_{A_1}\int_{A_2} C(y-x)\mathbf{1}_{\{x \leq 0 \leq y\}}\,\mu(dx)\mu(dy)$$

with C a normalizing constant, is an embedding of μ. Further, the process $B_{t\wedge T_S}$ is UI, and if μ has a second moment $\mathbb{E}(B^2_{T_S}) = \mathbb{E}T_S$.

Proof. C can be calculated, since μ is centred, by

$$\frac{1}{C} = -\int_{-\infty}^0 x\,\mu(dx) = \int_0^\infty y\,\mu(dy).$$

For $A \in \mathcal{B}([0,\infty))$ we can condition on X, Y to get

$$\mathbb{P}(B_{T_S} \in A) = \int_A \int_{-\infty}^0 \frac{-x}{y-x} C(y-x)\,\mu(dx)\mu(dy) = \mu(A)$$

and similarly for $A \in \mathcal{B}((-\infty, 0])$. So T_S is an embedding. Similarly, by conditioning on X, Y, we may calculate $\mathbb{E}T_S$ when μ has a second moment:

$$\begin{aligned} \mathbb{E}T &= \int_{-\infty}^{0}\int_{0}^{\infty} y|x|C(y-x)\,\mu(dx)\mu(dy) \\ &= \int_{-\infty}^{0} x^2\,\mu(dx) + \int_{0}^{\infty} y^2\,\mu(dy) \\ &= \mathbb{E}(B_{T_S}^2). \end{aligned}$$

Finally to deduce that $B_{t\wedge T_S}$ is UI we use Levy's upward martingale theorem (Rogers and Williams, 2000*a*)[Theorem II.69.5]. For the moment we suppose X, Y are $\mathcal{F}_0$-measurable and note that by the definition of T_S, $B_{t\wedge T_S} = \mathbb{E}(B_{T_S}|\mathcal{F}_t)$. Since $\mu \in \mathcal{L}^1$ the process is UI. $\square$

A.1.3 Wald's Lemma

Lemma A.3 (Wald's Lemma)**.** *If $(B_t)_{t\geq 0}$ is a Brownian motion with $B_0 = 0$ and T is a stopping time of the Brownian motion such that $\mathbb{E}T < \infty$ then*

(i) $\mathbb{E}B_T = 0$;

(ii) $\mathbb{E}B_T^2 = \mathbb{E}T$.

Proof. Fix $n \in \mathbb{N}$. Since $(B_t^2 - t)_{t\geq 0}$ is a martingale

$$\mathbb{E}(B_S^2) = \mathbb{E}S \leq \mathbb{E}(T\wedge n) \tag{A.1}$$

for all stopping times $S \leq T\wedge n$. Then $\sup_{S\leq T\wedge n}\mathbb{E}(B_S^2) \leq \mathbb{E}T$ and by Doob's $\mathcal{L}^2$-inequality $\mathbb{E}((B_{T\wedge n}^*)^2) \leq 4\mathbb{E}T$ where we write $B_t^* = \sup_{s\leq t}|B_s|$. We let $n \to \infty$ and deduce (by monotone convergence) that $\mathbb{E}((B_T^*)^2) < \infty$. So $B_{t\wedge T}$ is a $\mathcal{L}^2$-martingale and *(i)* holds.

By (A.1)

$$\mathbb{E}(B_{T\wedge n}^2) = \mathbb{E}(T\wedge n)$$

with the random variable on the left being dominated by ${B_T^*}^2 \in \mathcal{L}^1$. So we may take limits to deduce *(ii)*. $\square$

Bibliography

Azéma, J. and M. Yor (1979*a*), Une solution simple au problème de Skorokhod, *in* 'Séminaire de Probabilités, XIII (Univ. Strasbourg, Strasbourg, 1977/78)', Vol. 721 of *Lecture Notes in Math.*, Springer, Berlin, pp. 90–115.

Azéma, J. and M. Yor (1979*b*), Le problème de Skorokhod: compléments à "Une solution simple au problème de Skorokhod", *in* 'Séminaire de Probabilités, XIII (Univ. Strasbourg, Strasbourg, 1977/78)', Vol. 721 of *Lecture Notes in Math.*, Springer, Berlin, pp. 625–633.

Bass, R. F. (1983), Skorokhod imbedding via stochastic integrals, *in* 'Seminar on probability, XVII', Vol. 986 of *Lecture Notes in Math.*, Springer, Berlin, pp. 221–224.

Bertoin, J. and Y. Le Jan (1992), 'Representation of measures by balayage from a regular recurrent point', *Ann. Probab.* **20**(1), 538–548.

Blackwell, D. and L. E. Dubins (1963), 'A converse to the dominated convergence theorem', *Illinois J. Math.* **7**, 508–514.

Brown, H., D. Hobson and L. C. G. Rogers (2001*a*), 'The maximum maximum of a martingale constrained by an intermediate law', *Probab. Theory Related Fields* **119**(4), 558–578.

Brown, H., D. Hobson and L. C. G. Rogers (2001*b*), 'Robust hedging of barrier options', *Math. Finance* **11**(3), 285–314.

Chacon, R. V. (1977), 'Potential processes', *Trans. Amer. Math. Soc.* **226**, 39–58.

Chacon, R. V. and J. B. Walsh (1976), One-dimensional potential embedding, *in* 'Séminaire de Probabilités, X (Prèmiere partie, Univ. Strasbourg, Strasbourg, année universitaire 1974/1975)', Springer, Berlin, pp. 19–23. Lecture Notes in Math., Vol. 511.

Cox, A. M. G. and D. G. Hobson (2004*a*), Local martingales, bubbles and option prices. Preprint.

Cox, A. M. G. and D. G. Hobson (2004*b*), 'An optimal embedding for diffusions', *Stochastic Process. Appl.* **111**(1), 17–39.

Doob, J. L. (1984), *Classical potential theory and its probabilistic counterpart*, Vol. 262 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*, Springer-Verlag, New York.

Dubins, L. E. (1968), 'On a theorem of Skorohod', *Ann. Math. Statist.* **39**, 2094–2097.

Dubins, L. E. and D. Gilat (1978), 'On the distribution of maxima of martingales', *Proc. Amer. Math. Soc.* **68**(3), 337–338.

Dubins, L. E. and G. Schwarz (1988), 'A sharp inequality for sub-martingales and stopping-times', *Astérisque* (157-158), 129–145. Colloque Paul Lévy sur les Processus Stochastiques (Palaiseau, 1987).

Grandits, P. and N. Falkner (2000), 'Embedding in Brownian motion with drift and the Azéma-Yor construction', *Stochastic Process. Appl.* **85**(2), 249–254.

Hambly, B. M., G. Kersting and A. E. Kyprianou (2003), 'Law of the iterated logarithm for oscillating random walks conditioned to stay non-negative', *Stochastic Process. Appl.* **108**(2), 327–343.

Heath, D. (1974), Skorokhod stopping via potential theory, *in* 'Séminaire de Probabilités, VIII (Univ. Strasbourg, année universitaire 1972-1973)', Springer, Berlin, pp. 150–154. Lecture Notes in Math., Vol. 381.

Hobson, D. G. (1998*a*), The maximum maximum of a martingale, *in* 'Séminaire de Probabilités, XXXII', Vol. 1686 of *Lecture Notes in Math.*, Springer, Berlin, pp. 250–263.

Hobson, D. G. (1998*b*), 'Robust hedging of the lookback option', *Finance Stoch.* **2**(4), 329–347.

Hobson, D. G. and J. L. Pedersen (2002), 'The minimum maximum of a continuous martingale with given initial and terminal laws', *Ann. Probab.* **30**(2), 978–999.

Jacka, S. D. (1988), 'Doob's inequalities revisited: a maximal H^1-embedding', *Stochastic Process. Appl.* **29**(2), 281–290.

Madan, D. B. and M. Yor (2002), 'Making Markov martingales meet marginals: with explicit constructions', *Bernoulli* **8**(4), 509–536.

Meilijson, I. (1983), On the Azéma-Yor stopping time, *in* 'Seminar on probability, XVII', Vol. 986 of *Lecture Notes in Math.*, Springer, Berlin, pp. 225–226.

Meilijson, I. (2003), The time to a given drawdown in Brownian motion, *in* 'Seminar on probability, XXXVII', Vol. 1832 of *Lecture Notes in Math.*, Springer, Berlin, pp. 94–108.

Monroe, I. (1972), 'On embedding right continuous martingales in Brownian motion', *Ann. Math. Statist.* **43**, 1293–1311.

Monroe, I. (1978), 'Processes that can be embedded in Brownian motion', *Ann. Probability* **6**(1), 42–56.

Obłój, J. (2004*a*), The maximality principle revisited: on certain optimal stopping problems. Preprint `arxiv.org:math.PR/0405252`.

Obłój, J. (2004*b*), The skorokhod embedding problem and its offspring. Preprint `arxiv.org:math.PR/0401114`.

Obłój, J. and M. Yor (2004), 'An explicit Skorokhod embedding for the age of Brownian excursions and Azéma martingale', *Stochastic Process. Appl.* **110**(1), 83–110.

Pedersen, J. L. and G. Peskir (2001), 'The Azéma-Yor embedding in non-singular diffusions', *Stochastic Process. Appl.* **96**(2), 305–312.

Perkins, E. (1986), The Cereteli-Davis solution to the H^1-embedding problem and an optimal embedding in Brownian motion, *in* 'Seminar on stochastic processes, 1985 (Gainesville, Fla., 1985)', Birkhäuser Boston, Boston, MA, pp. 172–223.

Peskir, G. (1998), 'Optimal stopping of the maximum process: the maximality principle', *Ann. Probab.* **26**(4), 1614–1640.

Peskir, G. (1999), 'Designing options given the risk: the optimal Skorokhod-embedding problem', *Stochastic Process. Appl.* **81**(1), 25–38.

Pierre, M. (1980), Le problème de Skorokhod: une remarque sur la démonstration d'Azéma-Yor, *in* 'Seminar on Probability, XIV (Paris, 1978/1979) (French)', Vol. 784 of *Lecture Notes in Math.*, Springer, Berlin, pp. 392–396.

Revuz, D. and M. Yor (1999), *Continuous martingales and Brownian motion*, Vol. 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*, third edn, Springer-Verlag, Berlin.

Rogers, L. C. G. (1981), Williams' characterisation of the Brownian excursion law: proof and applications, *in* 'Seminar on Probability, XV (Univ. Strasbourg, Strasbourg, 1979/1980) (French)', Vol. 850 of *Lecture Notes in Math.*, Springer, Berlin, pp. 227–250.

Rogers, L. C. G. and D. Williams (2000*a*), *Diffusions, Markov processes, and martingales. Vol. 1*, Cambridge University Press, Cambridge. Foundations, Reprint of the second (1994) edition.

Rogers, L. C. G. and D. Williams (2000*b*), *Diffusions, Markov processes, and martingales. Vol. 2*, Cambridge University Press, Cambridge. Itô calculus, Reprint of the second (1994) edition.

Root, D. H. (1969), 'The existence of certain stopping times on Brownian motion', *Ann. Math. Statist.* **40**, 715–718.

Rost, H. (1971), 'The stopping distributions of a Markov Process', *Invent. Math.* **14**, 1–16.

Roynette, B., P. Vallois and M. Yor (2002), 'A solution to Skorokhod's embedding for linear Brownian motion and its local time', *Studia Sci. Math. Hungar.* **39**(1-2), 97–127.

Skorokhod, A. V. (1965), *Studies in the theory of random processes*, Translated from the Russian by Scripta Technica, Inc, Addison-Wesley Publishing Co., Inc., Reading, Mass.

Strassen, V. (1967), Almost sure behavior of sums of independent random variables and martingales, *in* 'Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66)', Univ. California Press, Berkeley, Calif., pp. Vol. II: Contributions to Probability Theory, Part 1, pp. 315–343.

Vallois, P. (1983), Le problème de Skorokhod sur $\mathbf{R}$: une approche avec le temps local, *in* 'Seminar on probability, XVII', Vol. 986 of *Lecture Notes in Math.*, Springer, Berlin, pp. 227–239.

Vallois, P. (1992), 'Quelques inégalités avec le temps local en zero du mouvement brownien', *Stochastic Process. Appl.* **41**(1), 117–155.